THEORY AND EXPERIMENTS OF ELECTRON-HOLE RECOMBINATION AT Si/SiO_2 INTERFACE TRAPS AND TUNNELING IN THIN OXIDE MOS TRANSISTORS

BY

JIN CAI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2000

LD
1780
2000
.C133

UNIVERSITY OF FLORIDA
3 1262 08554 4657

ACKNOWLEDGEMENTS

I am deeply indebted to Professor Chih-Tang Sah for his education, encouragement and patience throughout my graduate study at the University of Florida. I would like to thank Professors Arnost Neugroschel, Ewen Thomson and Sergei Obukhov for serving on my supervisory committee. I would also like to thank Professors Hunhui Sun and Xun Wang who led me into the field of solid-state physics during my senior year at Fudan University.

I am grateful to our industrial mentors from AMD, Harris, Intel, Motorola and Texas Instrument who provided the samples used in this dissertation research. I am also indebted to Dr. Tak Ning and Dr. Philip Wong for hiring me as a summer intern at IBM Research and Dr. Yuan Taur for his mentorship during my internship which exposed to me the latest state-of-the-art silicon technology. Additional thanks go to my previous and present colleagues, Dr. Michael Han, Dr. Guoxin Li and Yih Wang, for their professional helps and personal friendship. Financial support from Semiconductor Research Corporation and a Robert C. Pittman Doctoral Research Fellowship is gratefully acknowledged.

Finally I would like to thank my wife, Yijun Chen, and our parents, Fumin and Yumei Cai, Xuehong and Gendi Chen, for their continuous support throughout my five-year graduate education.

TABLE OF CONTENTS

ACKNOWLEDGEMENT ii

ABSTRACTS v

CHAPTERS

1 INTRODUCTION 1

2 INTERFACIAL ELECTRONIC TRAPS IN SURFACE CONTROLLED TRANSISTORS 6

2.1 Introduction 6
2.2 Carrier Recombination in MOS Transistor Channel Region 9
2.3 Carrier Recombination in MOS Transistor Junction Space Charge Region 18
2.4 Diffusion Limitation and Other Non-Ideal Factors 27
2.5 Application Examples 31
2.6 Summary 39

3 SPATIAL DENSITY PROFILE OF INTERFACE TRAPS FROM DCIV MEASUREMENTS 40

3.1 Introduction 40
3.2 Theoretical Analysis 41
3.3 Extraction of MOST Device Parameters 59
3.4 Profiling Interface Traps in Junction Space-Charge Regions 66
3.5 Interface Trap Profile in the Channel Region 70
3.6 Summary 80

4 GATE TUNNELING CURRENTS IN ULTRATHIN OXIDE MOS TRANSISTORS 81

4.1 Historical Survey 81
4.2 Recent Technology Motivations 86
4.3 Theory of Tunneling in MOS Structures 88
4.4 Theoretical Tunneling-Current Voltage Characteristics 97
4.5 Experimental Correlations 111

4.6 Summary 129

5 INTERBAND TUNNELING NEAR THE DRAIN/SUBSTRATE JUNCTION PERIMETERS IN MOS TRANSISTORS 131

5.1 Introduction 131
5.2 Theory of Drain Junction Perimeter Tunneling 132
5.3 Tunneling into Quantized Surface Accumulation Layer 139
5.4 Base-Terminal Current under Accumulation Gate Bias 148
5.5 Interface Trap Assisted Tunneling 158
5.6 Summary 170

6 SUMMARIES AND CONCLUSIONS 171

APPENDIX 175

REFERENCES 190

BIOGRAPHICAL SKETCH 200

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fullfilment of the
Requirement for the Degree of Doctor of Philosophy

THEORY AND EXPERIMENTS OF ELECTRON-HOLE RECOMBINATION AT Si/SiO_2 INTERFACE TRAPS AND TUNNELING IN THIN OXIDE MOS TRANSISTORS

by

Jin Cai

August 2000

Chairman: Chih-Tang Sah
Major Department: Electrical and Computer Engineering

Surface recombination and channel have dominated the electrical characteristics, performance and reliability of p/n junction diodes and transistors. This dissertation uses a sensitive direct-current current voltage (DCIV) method to measure base terminal currents (I_B) modulated by the gate bias (V_{GB}) and forward p/n junction bias (V_{PN}) in a MOS transistor (MOST). Base terminal currents originate from electron-hole recombination at Si/SiO_2 interface traps. Fundamental theories which relate DCIV characteristics to device and material parameters are presented. Three theory-based applications are demonstrated on both the unstressed as well as hot-carrier-stressed MOSTs: (1) determination of interface trap density and energy levels, (2) spatial profile of interface traps in the drain/base junction-space-charge region and in the channel

region, and (3) determination of gate oxide thickness and impurity doping concentrations. The results show that interface trap energy levels are discrete, which is consistent with those from silicon dangling bonds; in unstressed MOS transistors interface trap density in the channel region rises sharply towards source and drain, and after channel-hot-carrier stress, interface trap density increases mostly in the junction space-charge region.

As the gate oxide thins below 3nm, the gate oxide leakage current via quantum mechanical tunneling becomes significant. A gate oxide tunneling theory which refined the traditional WKB tunneling probability is developed for modeling tunneling currents at low electric fields through a trapezoidal SiO_2 barrier. Correlation with experimental data on thin oxide MOSTs reveals two new results: (1) hole tunneling dominates over electron tunneling in p+gate p-channel MOSTs, and (2) the small gate/drain overlap region passes higher tunneling currents than the channel region under depletion to flatband gate voltages. The good theory-experimental correlation enables the extraction of impurity doping concentrations, which complements the DCIV method.

Two fundamental theories of interband tunneling are developed to correlate with the V_{GB} dependence of drain/base p/n junction currents: (1) direct tunneling at the drain/base junction perimeter with and without the quantization effects in the base surface accumulation layer, and (2) interface trap assisted tunneling in the gate/drain overlap region. The second theory gives better correlation, which is further supported by the DCIV peaks originated from interface traps in the gate/drain overlap region.

CHAPTER 1
INTRODUCTION

One of the most significant advancements in silicon transistor technology development was the passivation of silicon surface by a thermally-grown oxide layer some forty years ago [1,2]. Today, the metal-oxide-semiconductor (MOS) transistor has become the most important building block of ultra-large-scale-integrated (ULSI) circuits. In MOS transistors, electrical currents conduct in silicon surface channels directly underneath the silicon/silicon dioxide interface. The mismatch of amorphous thermal SiO_2 and crystalline Si at the interface gives rise to point defects such as the unpaired silicon dangling bond which acts as an electronic trap. The intrinsic interface trap density after the thermal oxidation step can be as high as $5\times10^{12}cm^{-2}$, corresponding to a fractional occupancy of 0.6% of a (100) silicon plane, most of which are passivated by hydrogenation in a forming gas annealing step, to give a residual active interface trap density on the order of $10^{10}cm^{-2}$ in as-manufactured MOS transistors. However, during transistor operation, passivated interface traps can be reactivated and new interface traps can be generated which degrade MOS transistor performance by lowering its subthreshold slope and increasing its leakage current.

Traditional small-signal measurement techniques such as the MOS capacitance-voltage method can only resolve interface trap density higher than about $10^{11}cm^{-2}$ and can not detect manufacturing residual interface traps on state-of-the-art MOS transistors.

The seminal work on surface recombination and channel by Sah [3,4] was recently extended by Neugroschel and Sah [5] in 1995 to develop a sensitive technique known as the DCIV methodology that measures the DC recombination current of injected minority carriers at interface traps. The sensitivity is achieved by forward-biasing a p/n junction in a MOS transistor structure to exponentially raise the injected minority carrier concentration. In the past five years, many DCIV application papers and reports have been published which include delineation of interface trap generation/annealing kinetics on electrically-stressed transistors [6,7] and diagnosis of transistor design and manufacturing processes on pre-stress transistors [8-10]. This dissertation analyzes the theoretical basis of the DCIV method and develops theory-based new applications for characterization of scaled MOS transistors with very thin gate oxide.

In chapter 2, the gate-voltage controlled interface trap recombination current from the base-terminal is analyzed using the Shockley-Read-Hall steady-state recombination kinetics to provide analytical expressions for DCIV lineshape, linewidth, peak voltage and peak amplitude. It is predicted by the theory that DCIV peaks in the intrinsic to flat band gate voltage range originate from interface traps located in the channel area, while additional peaks in the surface accumulation range originate from interface traps covering the gated p/n junction space-charge region. A new application to determine interface trap density and energy level from the forward p/n junction bias dependence of the DCIV peak current is proposed and demonstrated on both unstressed and channel-hot-carrier stressed MOS transistors. The distortion of DCIV lineshape from minority carrier injection level and diffusion is also analyzed.

In chapter 3, spatial distribution of interface traps in the gated p/n junction space charge region (JSCR) is shown by theory to affect the DCIV lineshape in the gate voltage range between the weak inversion of the channel area and the weak inversion of the gate-overlapped drain area. In this gate voltage range, the recombination rate per unit interface trap is sharply peaked in the JSCR due to rapid lateral variation of minority carrier concentrations. Based on this discovery, an algorithm is developed to determine spatial density profile of interface traps which has a resolution of ~3nm. For interface traps located outside the JSCR in either the channel area or the drain area, the gate voltage that gives the peak recombination current depends only on transistor parameters such as the channel and drain doping concentrations. Thus these peak voltages provide a means to probe the MOS transistor design parameters. An analytical three-parameter formula is given to fit the forward junction bias dependence of DCIV peak voltage for the extraction of gate oxide thickness, impurity doping concentration and flatband voltage.

MOS transistor has been aggressively downscaled in the past 25 years to increase its speed while reducing its manufacturing cost at the same time. This trend will continue in the near future [11-12]. In each technology generation, gate oxide thickness is scaled almost in proportion to the channel length, while the Si/SiO_2 interfacial layer can not be scaled and becomes a more significant part of the total SiO_2 thickness in thin oxide MOS transistors. The DCIV method will continue to be a powerful tool for evaluating interface reliability as well as for characterizing transistor design. On thin oxide MOS transistors, DCIV measurements show base-terminal

current that rises with gate voltage, due to (1) increased gate oxide tunneling current at both inversion and accumulation gate voltages, and (2) increased current flow through forward-biased drain/base p/n junction at accumulation gate voltages. Quantitative understanding of these two phenomena will not only provide accurate baselines for DCIV peak analysis, but also enable new techniques for the characterization of scaled MOS transistors.

Recent experimental data on ultrathin oxide (1.0nm-3.0nm) MOS transistors [13] prompted us to develop a gate oxide tunneling theory that takes into account all three tunneling particle species (electrons, holes and valence electrons), which is presented in chapter 4. Theory-experimental correlation will prove two results: (1) hole is the dominant tunneling carrier in p+ gate p-channel MOS transistors, which was not recognized previously, and (2) the much smaller gate/drain overlap area passes higher current than the channel area in the gate voltage range between flatband and strong inversion [10,13]. Experimental tunneling currents from gate, source/drain, well and substrate terminals will be matched with the tunneling theory from which transistor design parameters such as oxide thickness, impurity doping concentrations in the channel and drain areas are determined.

In chapter 5, physical origins for the increased DCIV base-terminal current at accumulation gate voltages are investigated. Two theories of interband tunneling are considered: (1) direct tunneling at the drain/substrate junction perimeter including the effects of quantized surface accumulation layer, and (2) interface trap assisted tunneling in the gate-overlapped drain area. The second theory has a lower threshold gate voltage

and predicts a Shockley-Read-Hall recombination current peak from interface traps located in the overlap area, both of which are supported by experimental data.

We will summarize and conclude the dissertation in chapter 6.

CHAPTER 2
INTERFACIAL ELECTRONIC TRAPS IN SURFACE CONTROLLED TRANSISTORS

2.1 Introduction

Surface recombination and channel have dominated the electrical characteristics, performance, and reliability of p/n junction diodes and transistors. Due to the technological importance, extensive research efforts have been undertaken to study interfacial electronic traps at the Si/SiO_2 interface and to delineate their microscopic (atomic) origins [14]. Two kinds of density-of-states (DOS) spectra were repeatedly observed in the energy gap of silicon: two DOS peaks near the midgap and a U-shaped distribution that rises towards the two band edges. The discrete energy levels are characteristic of point defects at the Si/SiO_2 interface, while the U-shaped band-tail states are from random variations of Si-O bond length and bond angles [15]. These conclusions were supported by early works of molecular-orbital calculations of Si/SiO_2 interfacial electronic structures [16,17].

The peaked DOS were correlated to the trivalent silicon dangling bond (P_b center) on oxidized (111) silicon [18-20] and its two variations (P_{b0} and P_{b1} centers) on oxidized (100) silicon by the electron spin resonance (ESR) experiments [19,21]. Recent studies using ESR on oxidized silicon and spin-dependent recombination (SDR)

on MOS transistors reveal that the two P_b centers on (100) Si have different generation and annealing kinetics. Electric field stress including the non-uniform channel hot carrier stress [22] and the uniform Fowler-Nordheim stress [23] creates mostly P_{b0} centers, while both centers can be either passivated or generated by atomic hydrogen [24,25]. The structural difference among the three P_b centers can be grossly described by the direction of the single unpaired Si sp^3 hybrid orbital. It points to a normal <111> direction for both the P_b center on (111) Si and the P_{b0} center on (100) Si, while it points nearly along <211> for the P_{b1} on (100) Si [26]. Multiple DOS peaks of yet another origin were observed by 10.2eV light-induced hole injection in MOS capacitors annealed in oxygen ambient which could be from oxygen dangling bonds at the Si/SiO_2 interface [27].

While electron spin resonance has been used to identify the atomic structures of interface traps on large area (~1cm^2) oxidized silicon wafers, other techniques are more traditionally used to monitor interface traps in MOS transistors and MOS capacitors in manufacturing facilities. This includes MOS capacitance [28,29] and conductance [30] methods and more recently the charge-pump method [31] and DCIV method [5]. Routine manufacturing processes have reduced interface trap areal density to below $10^{10}cm^{-2}$ by slow cooling after high-temperature oxidation steps and by post-oxidation annealing in hydrogen. At this low density, the DCIV technique has its unique advantage in sensitivity over the other small-signal and transient techniques.

The principle of DCIV is the use of a surface potential controlling gate terminal voltage, V_{GB}, to modulate the base-terminal DC current, I_B, from electron-hole

recombination at the SiO_2/Si interface traps [3,4]. The DCIV methodology, which gives a family of I_B-V_{GB} curves from MOSTs and gated p/n junctions, was recently developed by Neugroschel and Sah [5] to monitor the density of oxide and interface traps generated by hot carriers (HC) in bipolar-junction and metal-oxide-field-effect transistors (BJTs and MOSTs) in order to determine the transistor failure rate and 10-year operation time-to-failure (TTF_{OP}) voltage [32]. Its application was extended to monitor the degradation rate of transistors stressed at high current-densities and low voltages (HJ), and to delineate the origin of hydrogen-related interconnect degradations [6,7]. Recent demonstrations showed that the DCIV method could serve as a highly sensitive diagnostic monitor for transistor designs [8] and manufacturing processes [8,33]. The DCIV lineshape (i.e. I_B vs V_{GB} around the $I_{B\text{-}peak}$) was also used to obtain the electronic and quantum-mechanical properties of the residual SiO_2/Si interface traps on state-of-the-art thermally grown thin gate oxides [34].

This chapter reports the elementary theory of the I_B-V_{GB} characteristics to help quantify the applications of the DCIV methodology for the extraction of fundamental and application-specific properties of transistors and interconnects and their materials, such as the physical (spatial location and density) and electronic (quantum density of states) properties of the residual and stress-generated interface and oxide traps, the dopant impurity concentration profiles, and the hydrogen sources. Analytical solutions and their physical models are presented first to illustrate the effects of material and interface trap parameters on the I_B-V_{GB} lineshape and on the $V_{GB\text{-}peak}$ value at the $I_{B\text{-}peak}$. Application examples are then given for the process-residue and hot-carrier-

generated interface traps on thin thermally oxidized gate-oxide of production MOS transistors.

2.1 Carrier Recombination in MOS Transistor Channel Region

In DCIV measurements, excess minority carriers are injected by a forward-biased p/n junction into the MOS-gated SiO_2/Si interface which covers MOST's channel region (denoted by MCR for mid-channel region) and drain-source junction space-charge region (JSCR). The p/n junction can be located either away from the gated interface, such as the well/substrate junction (bottom-emitter, BE) and a remote p/n junction (remote-emitter, RE) or under the gated interface, such as the source or drain p/n junction (source-emitter, SE, and drain-emitter, DE, or top-emitter with source-drain tied to the same forward bias voltage). The I_B-V_{GB} theory is presented for these injection bias configurations. The $I_{B\text{-}JSCR}$ current is especially important for factory applications to monitor channel-hot-carrier-generated interface traps, while the $I_{B\text{-}MCR}$ current is also useful for determining the electronic properties of the process-residual and stress-generated interface traps. Modifications of the theory by DCIV bias configurations, diffusion, dopant-impurity profiles and 2-dimensional geometries are then discussed.

Figure 2.1 is the energy band diagram of the metal-oxide-silicon structure with a forward-biased p/n+Si bottom emitter (BE) junction. It is labeled in detail to help describe the approximations of the analyses as follows. All voltages are normalized by kT/q. The BE is forward-biased at U_{BE} while the DC voltage applied to the gate

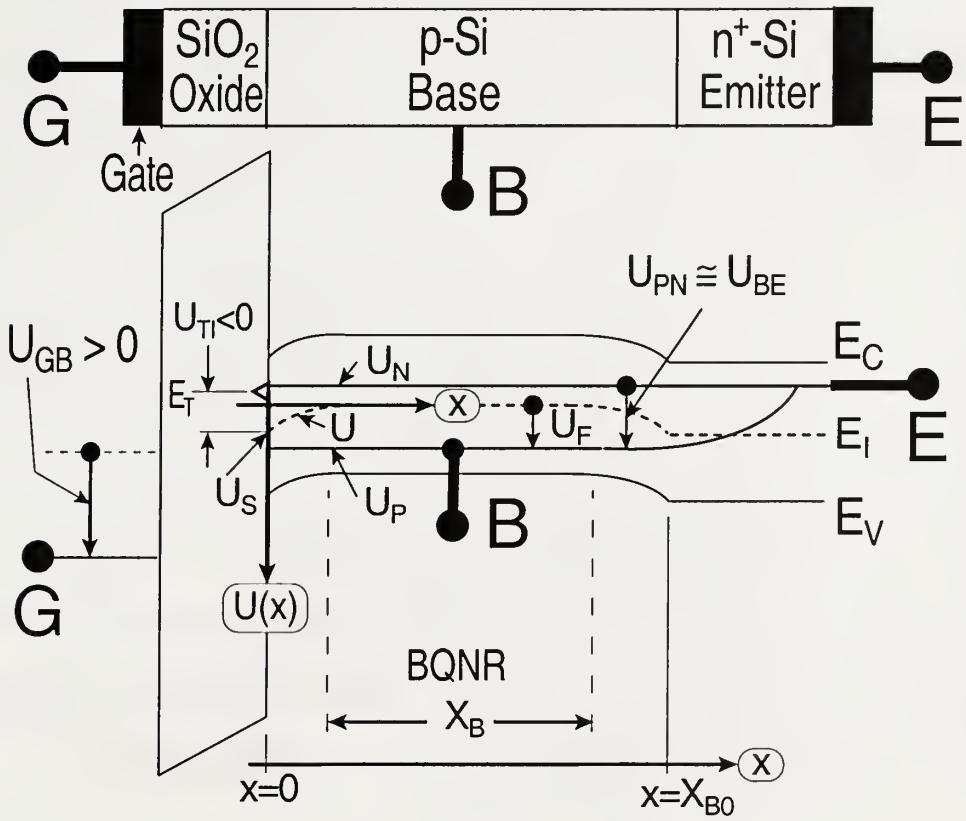

Fig. 2.1 Energy band diagram and cross-sectional view of a gated SiO_2/p-Si/n^+Si structure with the p/n^+ junction forward biased. U_N and U_P are respectively the electron and hole quasi-Fermi potentials normalized to thermal voltage (kT/q). Labeled on the figures are the terminologies (gate, base, and emitter), dimensions [X_{B0} and $X_B(V_{GB},V_{EB})$], and the normalized (to kT/q) voltages and quasi-Fermi-potentials (U_{GB}, U_{FB}=Flat-Band, and U_P, U_N).

relative to the p-Si base is U_{GB}. U_N and U_P are electron and hole quasi-Fermi potentials in p-Si and their difference, known as the quasi-Fermi-potential split, is $U_{PN}=U_P-U_N$ which is slightly smaller than the applied BE voltage, U_{BE}. A formula for ΔU_{PN} will be derived later. In the quasi-neutral region of p-Si base (BQNR), the quasi-Fermi-potentials are:

$$U_P = + \log_e\{[(N_{AA}^2+4n_i^2 \exp U_{PN})^{1/2}+N_{AA}]/2n_i\} \quad (2.1)$$

$$U_N = - \log_e\{[(N_{AA}^2+4n_i^2 \exp U_{PN})^{1/2}-N_{AA}]/2n_i\} \quad (2.2)$$

The electron and hole concentrations at the SiO_2/Si interface (N_S and P_S) are modulated by the gate voltage via bending the Si energy band. The total energy band bending is denoted by the surface potential $U_S=U(x=0,y)$.

There are four fundamental transition processes between the continuous band states of silicon crystal and the localized trap states with an energy level E_T in the Si energy gap, as shown in the transition energy band diagram of figure 2.2. The rate (event/second) of the four processes can be conveniently described by: (a) electron capture from the conduction band at $c_n n(N_{TT}-n_T)$, (b) electron emission to the conduction band at $e_n n_T$, (c) hole capture from the valence band at $c_p p n_T$, and (d) hole emission to the valence band at $e_p(N_{TT}-n_T)$. Here n and p are electron and hole concentrations in the conduction band and valence band respectively, N_{TT} is the total density and n_T is the electron-occupied density of trap states, and e's and c's are emission and capture rate coefficients of the four processes which depend on the energy levels of both the trap state and the band state. These transition processes are mostly thermal involving emission and capture of phonons ($\hbar\omega$). Since the indirect energy gap

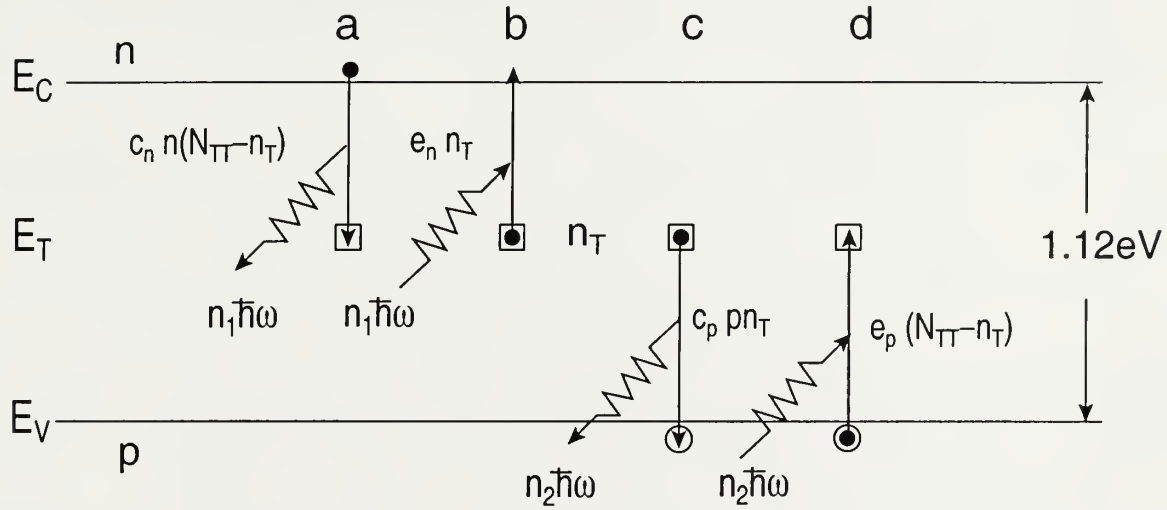

Fig. 2.2 A transition energy band diagram showing four fundamental transition processes between a band state and a gap state in silicon: capture of a condution band electron by the trap (a), emission of an electron to the conduction band from the trap (b), capture of a valence band hole by the trap (c) and emission of a hole to the valence band from the trap (d). The volume density of band electrons, band holes, electron-occupied traps and total traps are n, p, n_T and N_{TT} respectively. The rates of the four process are shown in terms of e's and c's. Purely thermal emission and capture processes involve multiple phonons.

of Si is 1.12eV and the maximum optical phonon energy in Si is only about 60meV, the thermal capture and emission processes could involve about ten phonons for mid-gap trap levels.

Theoretical calculation of a multi-phonon process was first done by Huang and Rhys [35] and later refined by Huang [36]. In contrast to the lattice scattering process which can be treated by taking the linear coupling of electron Coulomb potential and lattice vibration as the perturbation to calculate the scattering rate, the multi-phonon process can not be treated in such a way since the linear coupling term could only result in the emission and capture of a single phonon. Instead, the breakdown of the adiabatic approximation [37] was used as a perturbation to compute the transition rate between the band state and the trap state while the linear coupling term was used in the time-independent perturbation theory to compute the lattice distortion/relaxation [38] in the initial and final states. The first-principles calculation of the capture cross-sections in a multi-phonon process is rather laborious. For the purpose of developing a theory for the DCIV methodology, it would suffice to use the phenomelogical kinetic theory developed by Shockley and Read, and by Hall [39] which treats the fundamental capture and emission rates as constants independent of kinetic energies of band electrons and holes.

The steady-state areal rate, R_{SS}, of electron-hole recombination at a discrete-energy-level interface trap (the equilateral triangle in Fig. 2.1), located at U_{TI} from the intrinsic Fermi potential, with a density of N_{IT} (#/cm^2), is given by the Shockley-Read-Hall formula:

$$R_{SS} = \frac{(c_{ns}c_{ps}N_S P_S - e_{ns}e_{ps})}{c_{ns}N_S + e_{ns} + c_{ps}P_S + e_{ps}} \cdot N_{IT} \equiv R_{SS1} \cdot N_{IT} \qquad (2.3)$$

c_{ns}, c_{ps}, e_{ns} and e_{ps} are the electron-hole capture-emission rate coefficients at the interface traps. From detailed balance near thermal equilibrium, c's and e's are related by $e_{ns} = c_{ns}n_i\exp(U_{TI})$ and $e_{ps} = c_{ps}n_i\exp(-U_{TI})$. Using Boltzmann representation, $N_S=n_i\exp(U_S-U_N)$ and $P_S=n_i\exp(U_P-U_S)$, then (1) becomes

$$R_{SS} = \frac{(1/2)(c_{ns}c_{ps})^{1/2}n_i[\exp(U_{PN})-1]}{\exp(U_{PN}/2)\cosh(U_{S*}) + \cosh(U_{TI*})} \cdot N_{IT} \equiv R_{SS1} \cdot N_{IT} \qquad (2.4)$$

where $U_{S*}(x=0) = U_S+\log_e(c_{ns}/c_{ps})^{1/2}-(U_P+U_N)/2$, and $U_{TI*} = U_{TI} - \log_e(c_{ns}/c_{ps})^{1/2}$. The expression (2.4) is exact with no approximations other than the thermal Boltzmann distribution with lattice temperature T. It immediately shows the presence of a peak at $U_{S*}=0$. The peak magnitude and the surface potential at the peak are

$$R_{SS\text{-}peak} = \frac{(1/2)(c_{ns}c_{ps})^{1/2}n_i[\exp(U_{PN})-1]}{\exp(U_{PN}/2) + \cosh(U_{TI*})} \cdot N_{IT}$$

$$\equiv R_{SS1\text{-}peak} \cdot N_{IT} \qquad (2.5)$$

$$U_{S\text{-}peak} = (U_P+U_N)/2 - \log_e(c_{ns}/c_{ps})^{1/2} \qquad (2.6)$$

The peak formula (2.6), was derived by Sah et al. in 1957 [40]. The half width at half maximum (HWHM) of the R_{SS}-U_S line is:

$$\Delta U_S = \cosh^{-1}[2+\exp(-U_{PN}/2)\cosh(U_{TI*})] \qquad (2.7)$$

$$\approx \cosh^{-1}(2), \qquad \text{for } U_{PN} > 2(|U_{TI*}| + 4)$$

$$\approx U_{TI*} - (U_{PN}/2) \qquad \text{for } U_{PN} < 2(|U_{TI*}| - 4)$$

The HWHM drops to the narrowest, $(\cosh^{-1}2)*(kT/q)\approx 1.32kT/q=34mV$, at high V_{PN} and widens to U_{TI*} at low V_{PN}. Results (2.4)-(2.9) are valid at all injection levels in both p- and n-Si. For single-discrete level interface trap, they predict a peak body-terminal recombination current, given by $I_{B\text{-}peak} = qR_{SS1\text{-}peak}\int N_{IT}dydz$ that is proportional to $\exp(U_{PN}/n)$ with a transition of n from 1 to 2 over a small range of forward bias, about $\Delta U_{PN}=4$ or $\Delta V_{PN}\approx 100mV$. This result gives two important consequences. First, for a single species of interface traps with a spatially varying concentration, the DCIV lineshape will not change due to the y-distribution of N_{IT}. Second, if there are many interface trap species, each with a different energy level E_{TI*}, I_B is then just the superposition of the bell-shaped R_{SS} curves one from each of the energy levels or species with different peak locations in V_{GB}, resulting in a broadening and generally bell-shaped multi-peak DCIV curve [34] and an n-slope between 1 and 2 over a larger range of V_{PN}.

For low injection levels, traditionally defined as $N < N_{AA}/10$ in p-Si, we have $U_P=U_F>0$ for p-Si ($U_N=U_F<0$ for n-Si), where U_F is the majority carrier Fermi potential. This is the common application range of the DCIV methodology. At low levels, equation (4) becomes:

$$U_{S\text{-}peak} = U_F - U_{PN}/2 - \log_e(c_{ns}/c_{ps})^{1/2} \quad \text{(p-Si)} \qquad (2.8)$$

$$= U_F + U_{PN}/2 - \log_e(c_{ns}/c_{ps})^{1/2} \quad \text{(n-Si).} \qquad (2.9)$$

Thus, the peak is in the flat-band to intrinsic gate-voltage range ($0<U_S<U_F$, for p-Si). In this gate-voltage range, Sah [41] had given a formula for $V_S(V_{GB})$ listed in (2.10) below which we shall use to obtain the analytical solutions for the I_B-V_{GB} linewidth.

$$V_S = V_{GB} - V_{FB} - 2 \cdot \mathrm{sign}(V_S) \cdot V_{AA}\{[1+(V_{GB}-V_{FB}-kT\Delta_D/q)/V_{AA}]^{1/2} - 1\} \quad (2.10)$$

V_{FB} is the flatband voltage, $V_{AA}=\varepsilon_s q N_{AA}/2C_{OX}^2$ where $C_{OX}=\varepsilon_{ox}/X_{OX}$ is the oxide capacitance per unit area, and $\Delta_D \simeq 1$ in this range [41]. The last term in (2.10) is the voltage drop across the oxide layer, whose error is negligible when using $\Delta_D=1$ until V_S is about kT/q from the flatband condition. The V_{GB} at the peak can be solved from (2.10) in terms of $V_{S\text{-}peak}$ given by (2.8) and (2.9):

$$V_{GB\text{-}peak} - V_{FB} = V_{S\text{-}peak} + 2 \cdot \mathrm{sign}(V_{S\text{-}peak}) \cdot (V_{AA} \cdot |V_{S\text{-}peak}|)^{1/2} \quad (2.11)$$

Thus, the DCIV peak location is determined, via $V_{S\text{-}peak}$, by the substrate impurity concentration, trap parameters in terms of $\log_e(c_{ns}/c_{ps})$ and oxide thickness (from V_{AA}). The half-width at half maximum (HWHM) of the DCIV peak is:

$$\Delta V_{GB} = \Delta V_S + 2\sqrt{V_{AA}}\left[\sqrt{|V_{S\text{-}peak}|} - \sqrt{|V_{S\text{-}peak}|-\Delta V_S}\right] \quad \text{(flat-band side)} \quad (2.12)$$

$$= \Delta V_S + 2\sqrt{V_{AA}}\left[\sqrt{|V_{S\text{-}peak}|+\Delta V_S} - \sqrt{|V_{S\text{-}peak}|}\right] \quad \text{(intrinsic side)} \quad (2.13)$$

Since R_{SS} is symmetric around the peak in V_S, (2.12) and (2.13) show that the DCIV I_B-V_{GB} lineshape is asymmetric and slightly wider on the flat-band side of the peak than on the intrinsic side. The difference is on the order of $0.5(V_{AA}V_{S\text{-}peak})^{1/2}(\Delta V_S/V_{S\text{-}peak})^2$, which is more pronounced in transistors with thick oxide and high surface impurity concentration.

At high injection levels with $N > 10N_{AA}$, we have $U_P \simeq -U_N$ or the electron and hole concentrations are nearly equal in the BQNR, and the maximum surface recombination rate is near the flat-band. The exact result is $U_{S\text{-}peak} = \log_e(c_{ps}/c_{ns})^{1/2}$ derived from (2.6). So the peak could not move deeply into accumulation range even at extremely high forward bias since c_{ps}/c_{ns} is not likely to be outside of the range of 0.01 to 100 for any physically realistic bound state. The linewidth at the high injection level limit is obtained from the general MOS voltage equation in p-Si, $V_{GB}(V_S)$:

$$V_{GB}-V_{FB}-V_S = \mathrm{sign}(V_S)\cdot(kT/q)\cdot \{2U_{AA}[(e^{Us}-1-U_S)(a-1)+(e^{-Us}-1+U_S)(a+1)]\}^{1/2} \quad (2.14)$$

where $a = [1 + 4n_i^2\exp(U_{PN})/N_{AA}^2]^{1/2}$. At high injection levels, $a \approx 2\exp[(U_{PN}/2)-U_F] \gg 1$. Only one term in the { } of (2.14) will dominate, depending on the sign of U_S:

$$V_{GB}-V_{FB}-V_S \approx 2(kT/q)(U_{AA})^{1/2}\mathrm{sign}(U_S)\cdot \exp[(|U_S|+U_{PN}/2-U_F)/2] \quad (2.15)$$

The half-width of the DCIV line is then

$$\Delta V_{GB} - \Delta V_S = 2(kT/q)(e^{+\Delta Us/2}-1)[U_{AA}\sqrt{c_{ps}/c_{ns}}\exp(U_{PN}/2-U_F)]^{1/2} \quad (c_{ps}>c_{ns}, V_S>0) \quad (2.16)$$

$$2(kT/q)(1-e^{-\Delta Us/2})[U_{AA}\sqrt{c_{ns}/c_{ps}}\exp(U_{PN}/2-U_F)]^{1/2} \quad (c_{ps}<c_{ns}, V_S<0) \quad (2.17)$$

These results show that the Full Width Half Maximum (FWHM) has the $\exp(U_{PN}/4)$ dependence on forward bias at high injection levels. Thus, at low injection levels, the I_B-V_{GB} linewidth can be large which is determined by the trap level, E_{TI^*}. The

linewidth then decreases with increasing V_{PN} (to about FWHM=70mV for very thin oxide) until the onset of high injection level condition, beyond which it increases exponentially with V_{PN}.

Figure 2.3 shows the theoretical and normalized DCIV lineshapes or the unit steady-state recombination rate, R_{SS1} vs V_{GB}. The shape changes with increasing forward biases, V_{PN}, applied to a n+/p junction in p-Si with an acceptor impurity concentration of $N_{AA}=10^{17}cm^{-3}$, an oxide thickness of 5.0nm, and a discrete interface trap level at E_{TI*}=200meV. For $V_{PN}<E_{TI*}/q$, there is a flat-top region around the maximum R_{SS1}. The linewidth gradually decreases with increasing V_{PN} and reaches the high injection level asymptote as indicated by the dashed line at V_{PN}=900mV.

2.3 Carrier Recombination in MOS Transistor Junction Space-Charge Region

The preceding I_B-V_{GB} theory applies to interface traps over MOST's mid-channel region. This I_B component will be denoted as the $I_{B\text{-}MCR}$. The MCR formulas assumed that there is no lateral variation of the carrier concentrations, thus, the surface energy band bending U_S is constant, independent of y, at a given gate voltage. For interface traps over MOST's source and drain JSCR, the foregoing analysis needs to be extended to take into account of $U_S(y)$ in the JSCR. These interface traps contribute a second component to I_B, denoted as $I_{B\text{-}JSCR}$, when source, drain or both of them are forward biased. Figure 2.4 illustrates the cross-sectional view and the energy band diagram at the SiO_2/Si interface, x=0, of a forward-biased abrupt p/n+ junction. In Fig. 2.4, the build-in potential of the p/n+ junction is given by

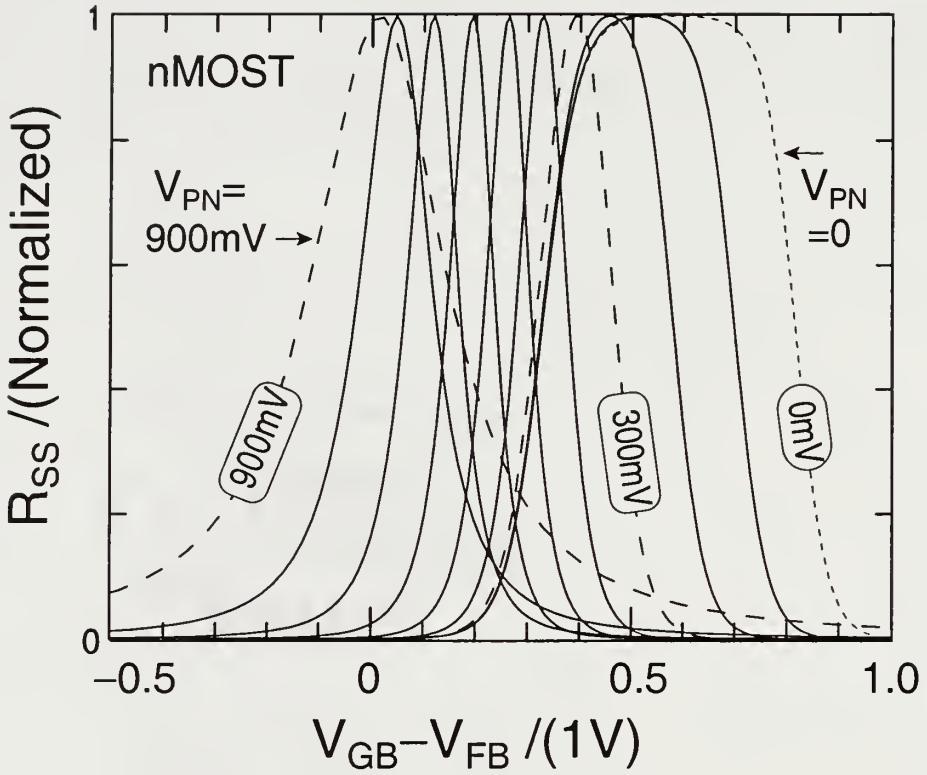

Fig. 2.3 Theoretical normalized DCIV lineshapes as a function of minority carrier injection levels, V_{PN}. Parameters used are as follows. V_{PN}=0-900mV at 100mV intervals with 0mV and 900mV the asymptotes. p-Si with N_{AA} = $10^{17}cm^{-3}$. $n_i=10^{10}cm^{-3}$. Oxide thickness X_{OX}=5.0nm. Discrete interface trap at E_{TI} = 200meV from the intrinsic Fermi level (~mid-gap) with $c_{ns}=c_{ps}=10^{-8}cm^3/s$, and N_{IT}=low→0 but arbitrary since normalized.

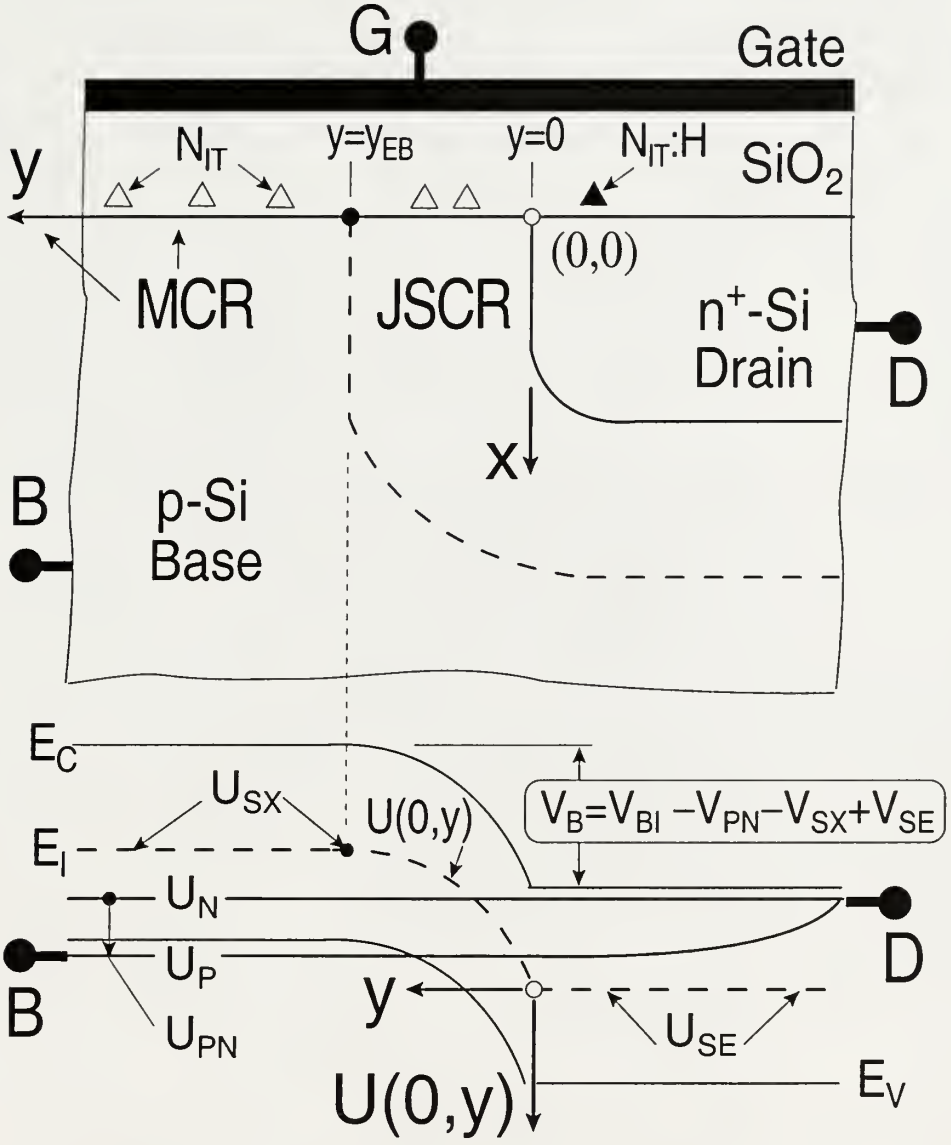

Fig. 2.4 The x-cross section view of the two-dimensional energy band diagram of a p/n^+ junction on the SiO_2/Si interfacial plane, x=0. The one-sided junction space charge region extends from y=0 to $y=Y_{EB}$ in the p-Si. V_{BI} = built-in potential of the p/n^+ junction, V_B = p/n+ barrier height = $V_{BI} - V_S - V_{PN}$, V_S = surface band bending outside the JSCR, and $V_{PN} = V_P - V_N$ = forward bias voltage applied to the p/n+ junction disregarding voltage or potential drops. All labeled potentials are normalized to (kT/q) with subscripted ($_{\#\#}$) symbol $U_{\#\#}=qV_{\#\#}/kT$.

$V_{BI}=(kT/q)\log_e(N_{AA}N_{DD+}/n_i^2)$. The total normalized barrier height at the surface of the p/n+ junction is:

$$U_B=U_{BI}-U_{PN}-U_{SX}+U_{SE} \qquad (2.18)$$

where U_{SX} and U_{SE} are the surface band bendings outside of the JSCR in the MCR and in the emitter (n+drain) region respectively. We use U(y) to represent the normalized surface potential in the JSCR along the interface from y=0 at the junction boundary to $y=Y_{EB}$ at the p-edge of the JSCR, with $U(Y_{EB})=U_{SX}$ given by boundary condition.

For the one-sided p/n+ junction, the n+ side JSCR is much smaller than the p-side JSCR and the lateral (y-direction) voltage drop across the n+ side JSCR can be neglected. However, the vertical (x-direction) surface band bending in the n+drain is significant at large negative gate biases even for highly degenerate doping concentrations. Figure 2.5 shows surface band bendings V_{SX} in part (a) and V_{SE} in part (b) for the gated p/n+ junction as a function of gate to p-substrate voltage V_{GB} with the drain to p-substrate forward bias $V_{DB}=-V_{PN}$ as a parameter. The results were obtained by solving one-dimensional MOS capacitor equations taking into account of non-equilibrium minority carrier concentrations. At $V_{GB}<-1V$ when the p-substrate surface is strongly accumulated, surface band bending in n+ drain with $N_{DD}=10^{19}cm^{-3}$ becomes more significant than that in the p-substrate with $N_{AA}=10^{17}cm^{-3}$. V_{SE} reaches −1.0V at $V_{GB}=-3V$ under a 100mV forward bias, while V_{SX} is at −0.25V at the same gate voltage and is independent of forward bias. It shows that under the surface accumulation V_{GB} range, V_{SX} is the dominant term in eqn. (2.18) for U_B and therefore indispensable in evaluating the surface potential U(y) in the JSCR.

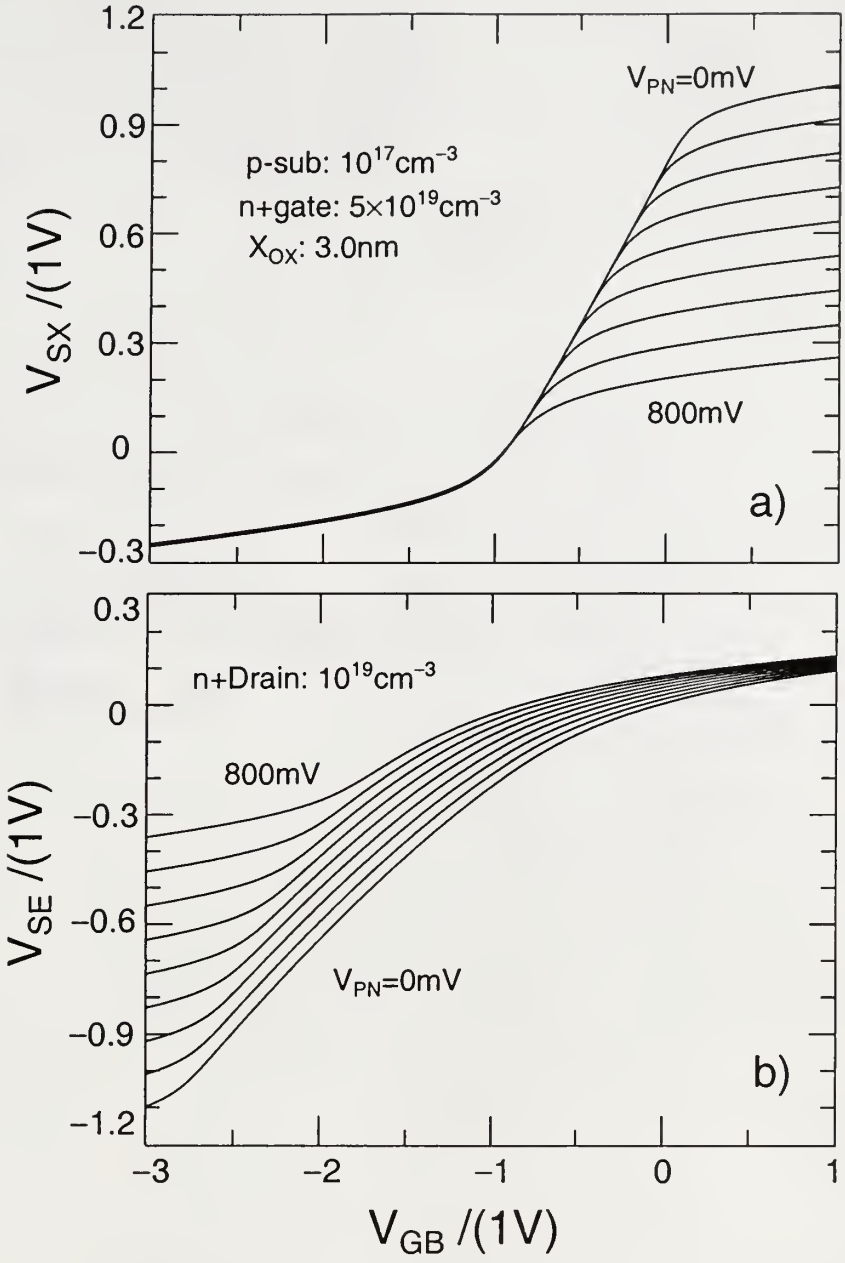

Fig. 2.5 Surface band bending in (a) p-substrate, V_{SX} and (b) n+ drain, V_{SE} as a function of gate voltage, with the forward bias on the drain/substrate junction as a parameter that varies from 0V to 800mV with a step size of 100mV. Transistor parameters include doping concentrations: $N_{AA}=10^{17}cm^{-3}$ in p-substrate, $N_{DD}=10^{19}cm^{-3}$ in n+ drain and $5\times10^{19}cm^{-3}$ in n+ polysilicon gate, and gate oxide thickness of 3.0nm.

The carrier concentration at $0<y<Y_{EB}$ in the JSCR are:

$$N_S(y) = n_i \exp[U(y)-U_N] = n_i \exp[U_{SX}+U(y)-U(Y_{EB})-U_N] \quad (2.19)$$

$$P_S(y) = n_i \exp[U_P-U(y)] = n_i \exp[U_P-U_{SX}-U(y)+U(Y_{EB})] \quad (2.20)$$

$I_{B\text{-}JSCR}$ is then obtained by integrating R_{SS}, given by (2.3) over the JSCR, which gives

$$I_{B\text{-}JSCR} = I_{BJ0} \int_0^{Y_{EB}} \frac{N_{IT}(y)\,dyZ}{\exp(U_{PN}/2)\cosh(U_{S*J}) + \cosh(U_{TI*})} \quad (2.21)$$

where for the p/n+ junction

$$I_{BJ0} = q(c_{ns}c_{ps})^{1/2}(n_i/2)[\exp(U_{PN})-1] \quad \text{(Ampere)} \quad (2.22)$$

$$U_{S*J} = U_{SX} + U(y) - U(Y_{EB}) - (U_P+U_N)/2 + \log_e(c_{ns}/c_{ps})^{1/2} \quad (2.23)$$

Two N_{IT} spatial distributions will be considered: (1) Localized $N_{IT}(y)=N_{IT}\delta(y-Y_T)$ and (2) uniform $N_{IT}(y)= N_{IT}$ = constant. The localized interface trap distribution is expected for channel-hot-carrier generated traps in MOSTs from Sah's Si:H bond breaking model [42] in which the generated silicon dangling bonds are concentrated at the location near the drain junction towards the source where the secondary hot carrier kinetic energy reaches the Si:H bond energy (3.1eV).

For the localized spatial distribution, $I_{B\text{-}JSCR}$ has a peak when $U_{S*J}=0$ in (2.23), which gives:

$$U_{SX\text{-}JSCR\text{-}pk} = (U_P+U_N)/2 - \log_e(c_{ns}/c_{ps})^{1/2} - [U(Y_T) - U(Y_{EB})]$$

$$= U_{SX\text{-}MCR\text{-}pk} - [U(Y_T) - U(Y_{EB})] \quad (2.24)$$

This shows a general result: the $I_{B\text{-}JSCR}$ peak is always at the accumulation gate-voltage side of the $I_{B\text{-}MCR}$ peak. For interface traps located closer to the p/n+drain junction boundary, y=0, $[U(Y_T\rightarrow 0)-U(Y_{EB})]$ in (2.24) is larger and the peak is further into the

accumulation V_{GB} range. Quantitative results can be obtained from analytical solutions when $|\partial E_x/\partial x| << |\partial E_y/\partial y|$ inside JSCR which we shall call the abrupt model, i.e. V_{GB} does not modulate the potential variation inside JSCR, only outside in the MCR. Then, using the depletion approximation, U(y) assumes the quadratic form illustrated in Fig.3 given by

$$U(y) - U(Y_{EB}) = (q/kT)(qN_{AA}/2\varepsilon_s) \cdot (Y_{EB} - y)^2$$

$$= (Y_{EB}-y)^2/2L_D^2 \qquad 0 \le y \le Y_{EB} \qquad (2.25)$$

where $L_D = [(kT/q)(\varepsilon_s/qN_{AA})]^{1/2}$ is the Debye screen length in p-substrate, $U(Y_{EB}) \equiv U_{SX}$ and the JSCR barrier height is given by $U_B = U(0) - U(Y_{EB}) = Y_{EB}^2/2L_D^2$. The thickness of the JSCR, Y_{EB}, is determined from the barrier height U_B. Using (2.18) and (2.25), we derive the following expression for the surface potential at the trap location $y=Y_T$ inside the JSCR in terms of surface band bendings outside of the JSCR, U_{SX} and U_{SE}:

$$U(Y_T) = U_{BI} - U_{PN} + U_{SE} -$$

$$2[U_{YT}(U_{BI}-U_{SX}-U_{PN}+U_{SE})]^{1/2} + U_{YT} \qquad (2.26)$$

where $U_{YT} \equiv Y_T^2/2L_D^2$. At $I_{B\text{-}JSCR}$ peak, $U(Y_T)$ satisfies equation (2.6), which together with (2.26) provide the following equation to relate the two surface band bendings at the peak:

$$U_{SX\text{-}JSCR\text{-}pk} = U_{BI}-U_{PN}+U_{SE\text{-}JSCR\text{-}pk}-(1/4U_{YT})\times$$

$$[U_{BI}-U_{PN}/2-U_F+U_{SE\text{-}JSCR\text{-}pk}+U_{YT}+\log_e(c_{ns}/c_{ps})^{1/2}]^2 \qquad (2.27)$$

The surface band bendings outside the JSCR in p-substrate and n+ drain are both determined by the gate voltage through the one-dimensional MOS capacitor equation.

Thus eqn. (2.27) provides a means to numerically solve the two surface band bendings at $I_{B\text{-}JSCR}$ peak.

The result of I_B-V_S HWHM (denoted by ΔV_S) in equation (2.7) applies to all N_{IT} locations, while the result of I_B-V_{GB} HWHM in (2.10) and (2.11) applies to N_{IT} in JSCR only if $U(Y_T)-U(Y_{EB}) < U_{SX\text{-}MCR\text{-}pk}$ so that $U_{SX\text{-}JSCR\text{-}pk}$ in (2.24) will not change sign, which corresponds to N_{IT} located near the p-edge of the JSCR. If the condition is not satisfied, then ΔV_S translates differently to ΔV_{GB} because the peak is now in the accumulation range while (2.10) of $V_{GB}(V_S)$ is valid only for the depletion range. To give an analytical solution in the accumulation range, we use the asymptotic $V_S(V_{GB})$ equation given by [41]:

$$U_S = -\log_e[(U_{GB}-U_{FB}-U_S)^2/4U_{AA} + \Delta_A] \qquad \text{(p-Si)} \quad (2.28)$$

with $\Delta_A \simeq 1$, then the HWHM of I_B-V_{GB} is given by

$$\Delta V_{GB} = \Delta V_S + 2(kT/q)\sqrt{U_{AA}}\exp(|U_{SX\text{-}JSCR\text{-}pk}|/2)[\exp(\Delta U_S/2)-1] \qquad \text{(accum. side)} \quad (2.29)$$

$$\Delta V_{GB} = \Delta V_S + 2(kT/q)\sqrt{U_{AA}}\exp(|U_{SX\text{-}JSCR\text{-}pk}|/2)[1-\exp(-\Delta U_S/2)] \qquad \text{(FB side)} \quad (2.30)$$

To compare the linewidth of JSCR peak to that of the MCR peak, we examine the change of surface potential $U(Y_T)$ with respect to an infinitesimal change of V_{GB} by taking the derivative of eqn. (2.26), which gives:

$$\begin{aligned}\Delta U(Y_T) &= [1-(U_{YT}/U_B)^{1/2}]\Delta U_{SE} + (U_{YT}/U_B)^{1/2}\Delta U_{SX} \\ &= [1-(Y_T/Y_{EB})^{1/2}]\Delta U_{SE} + (Y_T/Y_{EB})^{1/2}\Delta U_{SX} \qquad (2.31)\end{aligned}$$

In the limit that the interface traps are located near the MCR ($Y_T \rightarrow Y_{EB}$), the JSCR peak is close to the MCR peak with gate voltage corresponding to the weak inversion of p-

substrate surface where $|\Delta U_{SE}| << |\Delta U_{SX}|$. In the other limit that the interface traps are near the drain region ($Y_T \rightarrow 0$), the JSCR peak is close to the drain peak where $|\Delta U_{SE}| >> |\Delta U_{SX}|$, refer to Fig. 2.3. Therefore the two asymptotic limits of (2.31) are:

$$\Delta U(Y_T) \approx (Y_T/Y_{EB})^{1/2} \Delta U_{SX} \qquad Y_T \rightarrow Y_{EB}, \text{ near MCR} \qquad (2.32)$$

$$\Delta U(Y_T) \approx [1-(Y_T/Y_{EB})^{1/2}] \Delta U_{SE} \qquad Y_T \rightarrow 0, \text{ near drain} \qquad (2.33)$$

The above equations clearly show the DCIV linewidth dependence on the spatial location of the interface trap Y_T. If the interface traps are near the two edges of JSCR, the DCIV linewidths approach those of the MCR peak and the drain peak respectively, as $\Delta U(Y_T)=\Delta U_{SX}$ at $Y_T=Y_{EB}$ and $\Delta U(Y_T)=\Delta U_{SE}$ at $Y_T=0$. The DCIV linewidth increases as a function of $(Y_{EB}/Y_T)^{1/2}$ as the location of N_{IT} moves away from the MCR edge into the JSCR, it increases as a function of $1/[1-(Y_T/Y_{EB})1/2]$ as the NIT location moves away from the drain edge.

In case of a spatially constant N_{IT}, we shall evaluate the $I_{B\text{-}JSCR}$ integral, (2.21), analytically to describe the dependence of this current component as a function of V_{PN}. Since the integrand is expected to be a bell-shaped function with a maximum surface recombination rate at $y=Y_M$ that satisfies $U_{S*J}=0$ in (2.23) , we can use the Taylor series to expand the integrand about $y=Y_M$ and keep terms up to the second order. The result is then

$$I_B = \frac{2\sqrt{2}q\sqrt{c_{ns}c_{ps}}n_i N_{IT} Z}{3(q|E_M|/kT)} \cdot \frac{[\exp(U_{PN}) - 1]}{\exp(U_{PN}/4)\,[\exp(U_{PN}/2)+\cosh(U_{TI*})]^{1/2}} \qquad (2.34)$$

E_M is the electric field at the position $Y_M(V_{GB})$. E_M varies with a low-power-exponent of V_{PN} so that the I_B dependence on V_{PN} is dominated by the exponential dependence of V_{PN} shown in numerator and denominator of (20), which predicts a n=1.33 to n=2 transition from low forward biases to high forward biases with a transition range of about U_{PN}~8≈200mV.

2.4 Diffusion Limitation and Other Non-Ideal Factors

DCIV measurement on MOS transistors can use four different bias configurations to inject minority carriers to the SiO_2/Si interface [8], as illustrated in figure 2.6. They are the drain-emitter (DE), source-emitter (SE), top-emitter (TE) with both source and drain junctions tied and forward-biased together, and the bottom-emitter (BE) in transistor structures having a well/substrate p/n junction which can be forward biased. In each of the bias configurations, the p/n junctions not forward biased are zero-biased (they can also be reverse-biased or even forward-biased at a lower voltage). In the first three gated-emitter configurations (DE, SE, TE), the body-terminal current I_B has two surface recombination components: from the MCR, $I_{B\text{-}MCR}$, and from the one or two emitter-base JSCRs, $I_{B\text{-}JCR}$, while in the ungated BE configuration, I_B contains only $I_{B\text{-}MCR}$ because the drain and source p/n junction collectors are grounded so their $I_{B\text{-}JSCR}$=0. However, the zero-biased source and/or drain p/n JSCR's also reduce the interface minority carrier concentrations in the MCR. This lowers $I_{B\text{-}MCR}$ and changes the I_B-V_{GB} lineshape which becomes appreciable in short channels. A more prominent

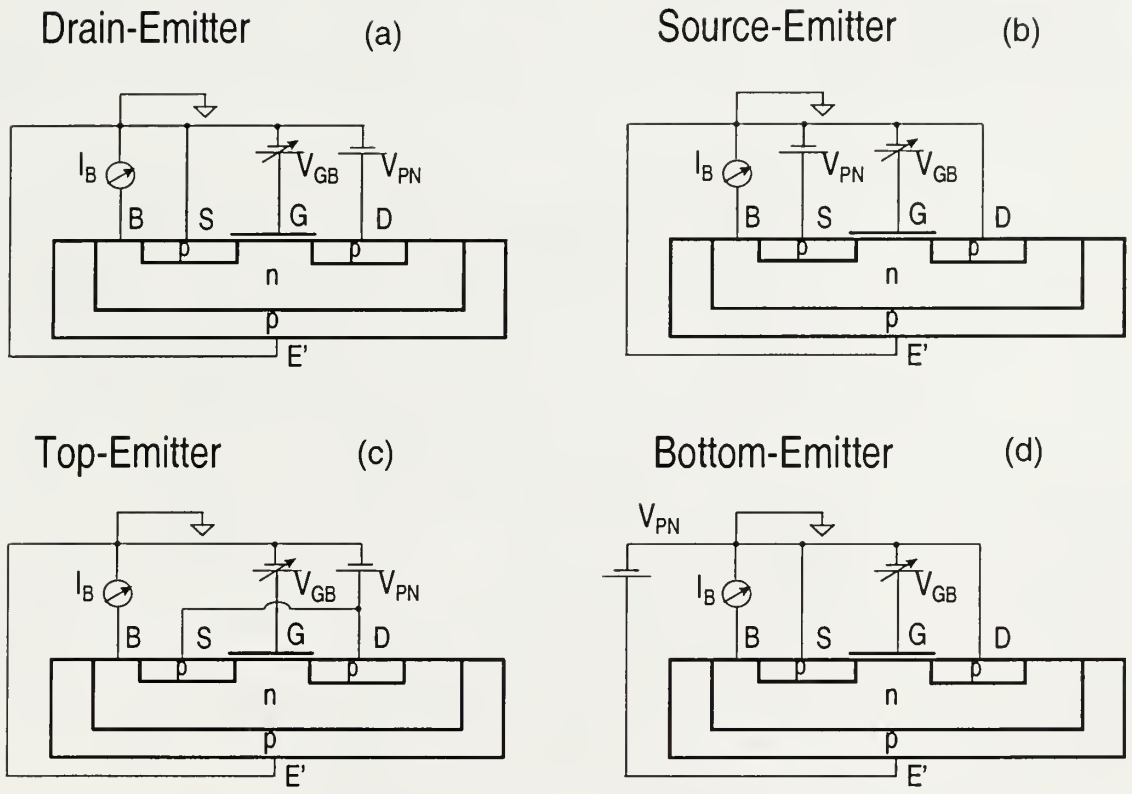

Fig. 2.6 Four DCIV bias configurations: (a) Drain-emitter, (b) Source-emitter, (c) Top-emitter, and (d) Bottom-emitter.

lowering and shape-change of $I_{B\text{-}MCR}$ arises from the 2-dimensional geometrical edge effect due to the proximity of the grounded collector and forward-biased emitters in the thin-base whose base thickness under the drain and source junction, is comparable with the metallurgical channel length. Another source of deviation from the simple ideal theory comes from the 2D impurity doping profiles as a result of channel and junction engineering. Minority carriers injected from the body/well junction in BE-DCIV will have a drift component due to the built-in field $E_X(x) = (kT/q)d(\log_e N_{AA})/dx$ from the vertical impurity profile. The vertical profile also gives a variation of the injected minority carrier concentration near the source and drain junction perimeters. These would give a difference between the TE and BE lineshapes. The channel regions that have different $N_{AA}(x=0,y)$ and different oxide charge density, $Q_{OT}(y)$, will shift V_{GB} of the DCIV peaks which are combined into a broader peak.

Diffusion of the injected carriers in the BQNR (Base Quasi Neutral Region) can affect and even limit the terminal I_B, especially the I_B-V_{GB} lineshape. For example, the interface recombination rate could be so high that the minority carrier concentration at the SiO_2/Si interface is decreased below the injected value at the injecting p/n junction boundary, $N_S < N_0 = (n_i^2/N_{AA})\exp(U_{BE})$. In the limit of $N_{IT} \rightarrow \infty$, $N_S = (n_i^2/N_{AA})$ or $U_{PN}(x=0)=0$. The interface recombination and base diffusion mechanisms are in series. In contrast, bulk-JSCR recombination (the Sah-Noyce-Shockley or SNS current component) and base diffusion mechanisms are in parallel [40]. Thus, the gate-controlled terminal I_B in the DCIV methodology is limited by the smaller of the two series current mechanisms, base diffusion current at low V_{PN} and interface

recombination current at higher V_{PN}. A simple analytical solution to include diffusion limitation is obtained using $V_{BE} = V_{PN}+\Delta V_{PN}$ and assuming that ΔV_{PN} is mainly in the BQNR since the surface-space-charge layer is much thinner than the BQNR. Then, $\Delta V_{PN}(V_{BE})$ at the $I_{B\text{-}peak}$ can be obtained as follows by equating the interface recombination current to the diffusion current:

$$J_{B\text{-}peak} = [qn_i(c_{ns}c_{ps})^{1/2}N_{IT}/2]\cdot [\exp(U_{PN})-1]/[\exp(U_{PN}/2)+\cosh(U_{TI*})] \quad (2.35)$$

$$= J_D \simeq (qD_B/X_B)(n_i^2/N_{AA})\exp(U_{PN})[\exp(\Delta U_{PN})-1] \quad (2.36)$$

which gives

$$\exp(\Delta U_{PN}) - 1 = (N_{AA}X_B/n_iD_B)[(c_{ns}c_{ps})^{1/2}N_{IT}/2]\cdot [1-\exp(-U_{PN})]/[\exp(U_{PN}/2)+\cosh(U_{TI*})] \quad (2.37)$$

$$= 1000\cdot[1-\exp(-U_{PN})]/[\exp(U_{PN}/2)+\cosh(U_{TI*})] \quad (2.38)$$

$$\simeq 1000\cdot/\cosh(U_{TI*}),$$

$$\text{for } 0<U_{PN}<2[\log_e\cosh(U_{TI*})-2.3]\simeq 2(|U_{TI*}|-3.0) \quad (2.39)$$

$$\simeq 1000\cdot\exp(-U_{PN}/2),$$

$$\text{for } U_{PN}>2[\log_e\cosh(U_{TI*})+2.3]\simeq 2(|U_{TI*}|+1.6) \quad (2.40)$$

D_B is the minority carrier diffusivity in the base well of X_B thick. The value of 1000 is obtained from using the following estimated typical values: $c_{ns}=c_{ps}=10^{-8}cm^3/s$ for a 0.3nm diameter neutral potential well of the interface traps and $N_{IT}=2x10^9cm^{-2}$ (which give a surface recombination velocity of $S=(c_{ns}c_{ps})^{1/2}N_{IT}=20cm/s$); $N_{AA}=10^{17}cm^{-3}$ at

$x=X_B$, $n_i=10^{10}cm^{-3}$, $D_B=10cm^2/s$, and $X_B=10^{-4}cm$ which gives $(N_{AA}X_B/2D_Bn_i)=50s/cm$.

It is evident from the approximate results given in (2.39) and (2.40) that there could be a significant ΔU_{PN} only at low U_{PN}'s and only if the interface trap energy level is near the mid-gap. The DCIV lineshape in this case will be distorted and have a diffusion-limited flat top. An interface-recombination-limited threshold V_{PN} can be defined by equating the asymptotic current equations limited by recombination (2.36) and diffusion (2.37). Using the numerical values just listed, this gives

$$U_{PN-th} = 2[6.908 - \cosh(U_{TI*})] \qquad (2.41)$$

The diffusion current J_D given by (2.36) for constant N_{AA} can be generalized to graded and retrograded profiles, $N_{AA}(x)$, by replacing the pre-exponential factor with the exact integral given by $\int [N_{AA}(x)/D_B(x)n_i^2(x)]dx$ which is integrated over the quasi-neutral base, x=0 to $x=X_B$. This is the well-known Gummel Number.

2.5 Application Examples

DCIV measurements have been performed extensively since its revival in 1995 [5] on silicon micron and submicron MOSTs manufactured by 1988-1998 technologies with channel aspect ratios of W/L=100μm/100μm to 10μm/0.1μm and gate oxide thicknesses from 30nm to 1nm. This section will present the DCIV curves obtained on both unstressed and hot-carrier-stressed MOSTs to illustrate the theoretical analyses of the experiments. The data were from the large pMOSTs to minimize the 2-D effects.

The pMOSTs have W/L=50μm/50μm, 10.5nm gate oxide, N_{DD}(interface) ≈ $2x10^{17}cm^{-3}$ and a graded n-well with p+buried bottom emitter.

The first examples are from unstressed pMOSTs where the I_B-V_{GB} peak originates from the residue interface traps of the manufacturing processing steps. Figure 2.7(a) shows a family of normalized experimental BE-DCIV curves in solid lines and theoretical curves in dashed lines based on E_{TI*}≈228meV and N_{IT}≈$7.8x10^{9}cm^{-2}$. These numerical values were obtained from a 2-parameter least-squares-fit [33] of the $I_{B\text{-}peak}$ vs V_{PN} shown in Fig. 2.7(b) which shall be coined as the DCIV Peak-Plot (P-Plot). The P-plot fits very well only the discrete-trap-level theory while the data deviate substantially from the theory computed from constant or U-shaped interface trap energy level distributions. Diffusion current was not included in this fit. Fig.7(a) indicates a good theoretical account of the lineshape at the strong inversion side while the experimental halfwidth is wider on the flatband side. Several sources contribute to the difference including: diffusion limitation at these low V_{PN} values which would broaden the linewidth, larger surface impurity concentration than assumed from C-V measurements and graded vertical doping profiles.

Figure 2.8 shows the differences among the DE-, SE-, TE- and BE-DCIV curves measured at V_{PN}=400mV on the same unstressed pMOST. The three gated-emitter configurations (DE, SE and TE) all show one narrow peak indicating low N_{IT} concentrations inside the gated JSCR. Overlap of SE and DE curves shows good source/drain symmetry. The different $V_{GB\text{-}peak}$ locations between TE and DE or TE and SE come from the lateral variation of the minority carrier (holes) concentration at the

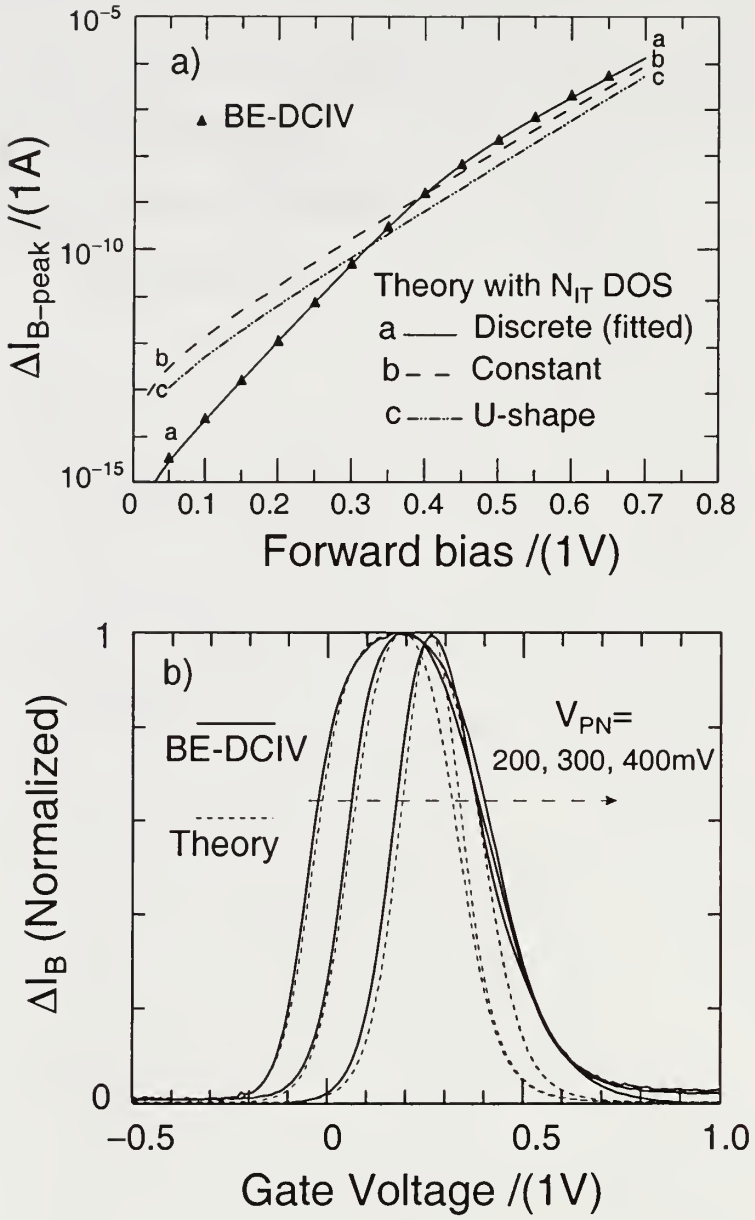

Fig. 2.7 Experimental BE-DCIV's and Peak-Plots on unstressed 50/50 pMOST compared with theories. Parameter values are: E_{TI*}=228meV and N_{IT}=7.8x10^{9}cm^{-2} from least-squares-fit to the discrete-level theory using c_{ns}=c_{ps}=10^{-8}cm^3/s (assumed), n_i=10^{10}cm^{-3}, and from C_g-V_{GB} data, N_{AA}(x=0)=2x10^{17}cm^{-3}, and X_{ox}=10.5nm. **(a)** Peak Plots ($I_{B\text{-peak}}$ vs V_{PN}) of experimental data (points) and theory from three assumed DOS's of interface traps: a: Discrete-level given above, b: constant DOS over $E_{G\text{-}Si}$ with D_{IT} = 7.1x10^9/cm^2-eV, and c: U-shaped D_{IT} = 2.5x10^{9}cosh(E_{TI}/200meV)/cm^2-eV. **(b)** Lineshape comparison with one-discrete-level theory.

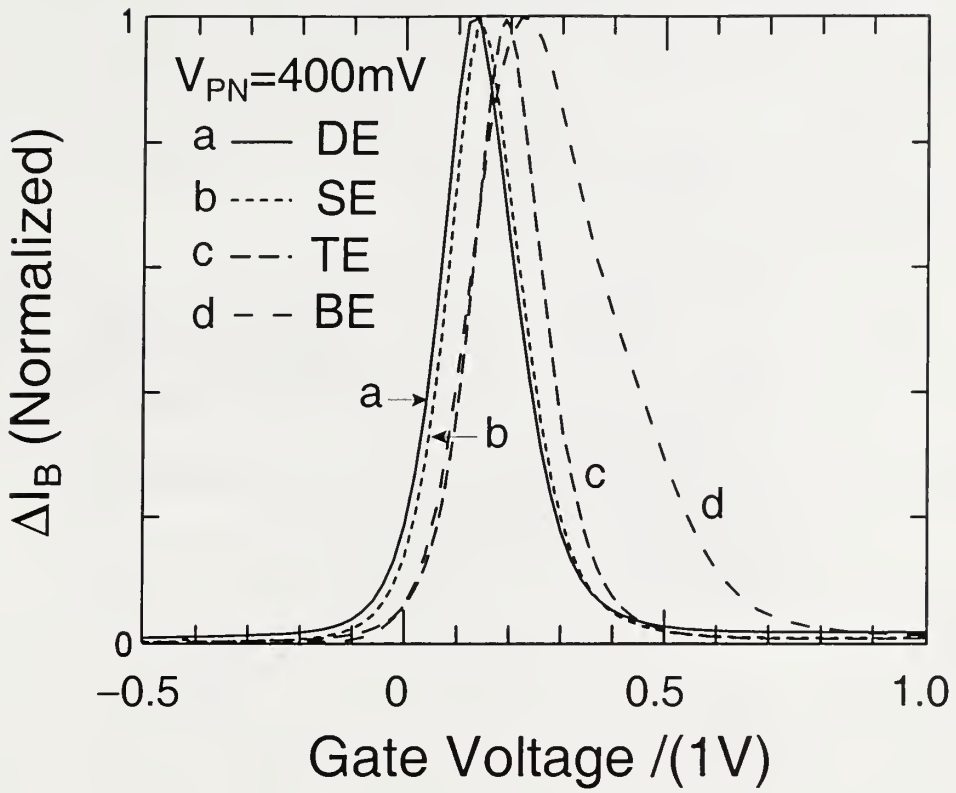

Fig. 2.8 Experimental lineshape comparison of DCIV at V_{PN}=400mV from four different measurement configurations (DE, SE, TE, and BE) on the 50/50 pMOST.

interface, $P_S(y)$, from $P_S(0)=0$ at the shorted source junction to $P_S(L)=P_{SS}[\exp(U_{PN})-1]$ at the forward-biased drain junction in the DE-DCIV bias configurations or vise versa in the SE-DCIV bias configuration. $P_S(y)$ is further reduced due to the finite diffusion length towards the MCR from the injection edge of the drain or/and source junction. These give a smaller effective channel interface recombination area and also a smaller effective V_{PN} which would shift the peak location towards the $-V_{GB}$ side as indicated by (2.9). Figure 2.8 shows that the I_B-V_{GB} linewidth of the ungated emitter (BE) is almost doubled in comparison with those of the gated emitter (SE, DE and TE), which is due to carrier injection against a diffusion barrier from graded BQNR. These effects on the lineshape from junction and geometrical edge effects and impurity concentration profile, indeed, provide flexible variables that can be leveraged in the monitor application of the DCIV methodology for optimization of transistor designs and fabrication processes.

For the second sets of application examples, the DCIV theory is used to evaluate the interface traps generated during the substrate-hot-carrier (SHC) and channel-hot-carrier (CHC) stresses of the same 50/50 pMOSTs used in the preceding unstressed examples. During the SHC stress, the p+buried/n-well junction is forward biased to inject a high concentration of holes into the strongly inverted hole surface channel which is reverse biased by a high reverse voltage applied simultaneously to the drain and source p+/n-well junctions. This high reverse voltage accelerates the injected holes to a kinetic energy greater than the Si:H interfacial bond energy (3.1eV) in order to release the hydrogen and generate the silicon dangle bonds, Si•, which are the interface traps. Reasonably good lateral uniformity of interface trap generation was expected

from using low-enough V_{BE} and strong-enough inversion V_{GB} to minimize the lateral drop of the vertical acceleration potential due to channel ohmic resistance. Figure 2.9(a) shows the I_B-V_{GB} DCIV lineshapes which indicates a larger experimental linewidth suggesting diffusion limitation. Figure 2.9(b) shows the DCIV Peak-Plot, $I_{B\text{-}peak}$-V_{PN}, which includes also the unstressed data that coincide the stressed data for V_{PN}<400mV further indicating diffusion limitation. The stress-generated interface traps are still quite discrete as indicated by the good fit to the discrete-energy-level model at high V_{PN}'s. The least-squares-fit (LSF) stressed-generated trap energy level, E_{TI*} = 280meV, should shift towards midgap when diffusion is taken into account.

A second example of stress-generated interface traps was obtained using the channel-hot-carrier stress on the same 50/50 pMOST whose DCIV at V_{PN}=350mV and P-Plot are shown in Fig. 2.10. The unstressed peak near the intrinsic gate voltage (V_{BG}≃+0.1V) does not grow at all which is consistent with its MCR location where there are few if any hot carriers. The growing post-stress peak is broad (FWHM ~ 2V) and in the accumulation range (V_{GB}=+1.8V). It has an n→2 over many decades in the P-Plot (inset of Fig.8) which suggests that the interface traps are mainly generated in the drain JSCR, as anticipated by the JSCR-DCIV theory just described, and the CHC stress theory [42] where the channel hot holes and the secondary hot electrons are generated in the drain junction space charge region in the pMOST during stress.

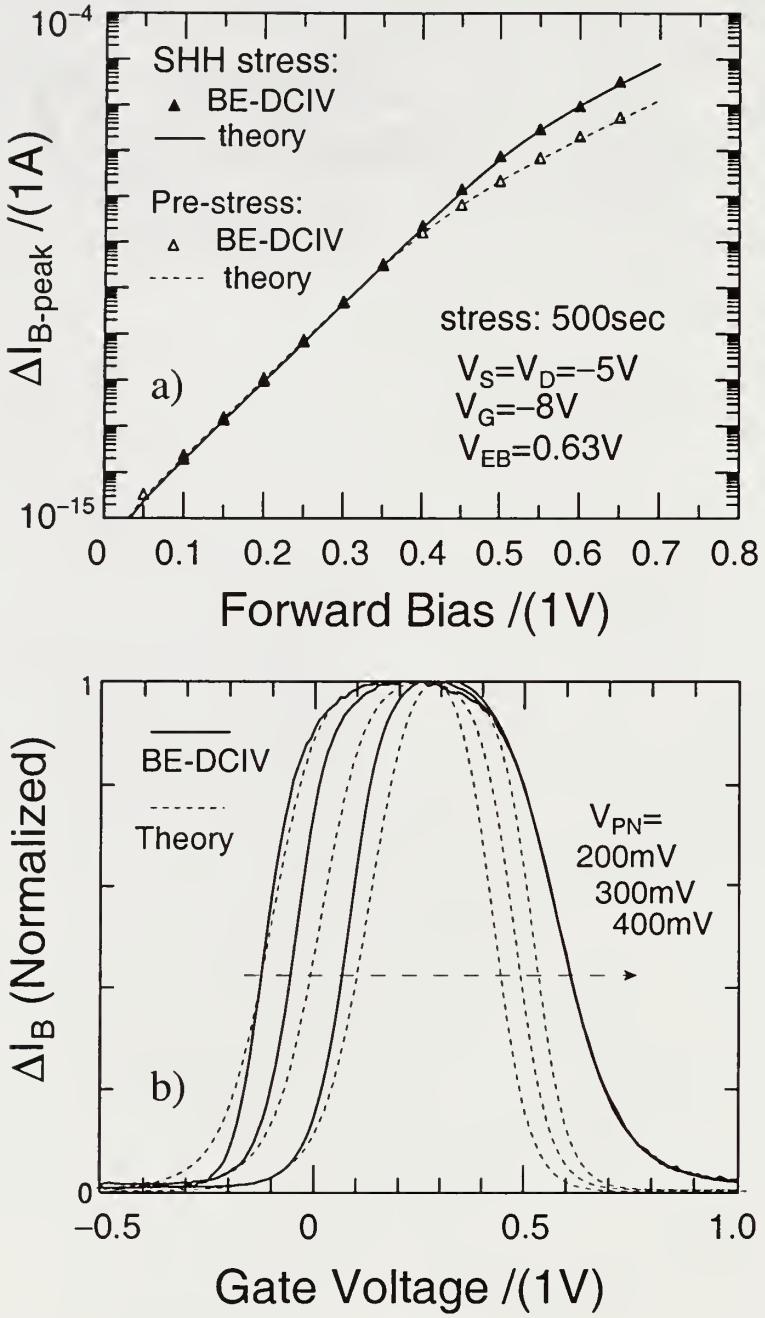

Fig. 2.9 Experimental and theoretical LSFitted BE-DCIV's from substrate-hot-carrier (hole) stressed 50/50 pMOST. **(a)** The Peak-Plots $I_{B\text{-peak}}$ vs V_{PN} showing a discrete-energy-level interface trap at E_{TI*}=280mV and $N_{IT}=5\times10^{10}cm^{-2}$. **(b)** Lineshape comparison with the discrete-level theory. Other parameter values are the same as those of Fig. 2.7.

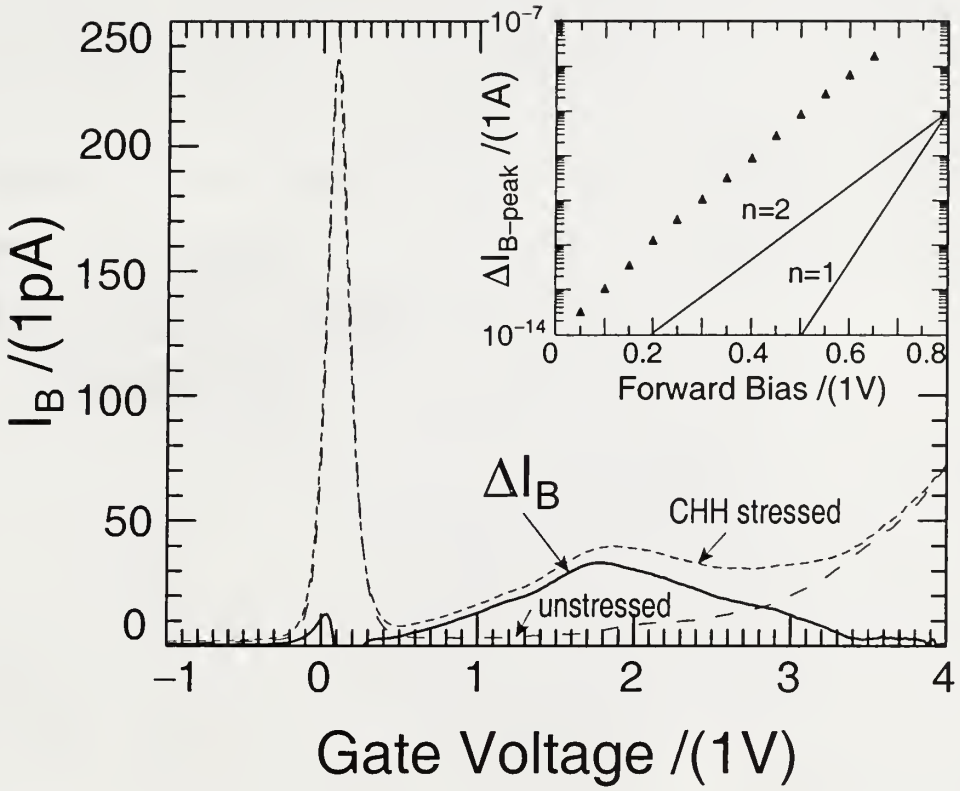

Fig. 2.10 Experimental DE-DCIV's taken on a channel-hot-carrier (hole) stressed pMOST, with the Peak-Plot of the stressed-generated $\Delta I_{B\text{-peak}}$ in the inset. CHC (CHH) stress condition: V_{SB}=0V, V_{DB}=−7.0V, V_{GB}=−1.0V and t=1000 sec.

2.6 Summary

Theory and experiments on gate-voltage modulation of the recombination rate at the interfacial electronic traps are quantitatively correlated in prestress and two poststress (channel and substrate hot carrier stresses) DCIV applications. The wider experimental linewidths could be an indication of diffusion limitation in the low forward bias range, and from 2-dimensional effects at high forward bias range due to lateral variations of the interface minority carrier concentration along the gated conduction channel. The examples illustrate the identification of discrete interface trap energy levels and the near midgap energy level positions.

CHAPTER 3
SPATIAL DENSITY PROFILE OF INTERFACE TRAPS FROM DCIV MEASUREMENTS

3.1 Introduction

The DC basewell terminal (or pad) current, I_B, modulated by the applied gate/base DC voltage, V_{GB}, in MOS transistors was a method recently reactivated to monitor electric-field-stress generated interface traps as a transistor reliability monitor [5] and to serve as a pre-stress diagnostic monitor for transistor design and processing [8]. The V_{GB}-modulated I_B arises from recombination of the injected minority carriers by a forward-biased p/n junction (Drain/Base, Source/Base, or Substrate/Basewell) with the majority carriers in the basewell, at the SiO_2/Si interface traps under the gate oxide. Electron-hole recombination can occur at interface traps under the gate oxide in three regions [3,4]: (1) the mid-channel region MCR (2) the drain and source junction space-charge regions D-JSCR and S-JSCR, and (3) drain and source lowly-doped extension regions D-LDER and S-LDER. A detailed theoretical analysis was presented in chapter 2 for the MCR and JSCRs. Some of our recent application demonstrations [8,33,34] focused on interface traps N_{IT} in the MCR. However, recombination at interface traps in the other two regions, JSCR and LDERs, becomes increasingly important in unstressed transistors as the channel length is scaled down and it is well-known that recombination in JSCR dominates in stressed transistors [8,43,44]

regardless of the channel length [8,33]. It was anticipated [8] and analytically demonstrated in the previous chapter that the I_B-V_{GB} lineshape and its V_{PN} dependencies are affected by two lateral variations: the interface traps, $N_{IT}(y)$, [8] and the interface concentration of the injected minority carriers N(x=0,y) [8] in n-channel MOSTs or P(x=0,y) in p-channel MOSTs. This spatial dependency was analytically shown in chapter 3 to be substantial in the JSCR due to the rapid spatial variation of the carrier concentrations in the JSCR. We will now describe the procedures for obtaining the $N_{IT}(y)$ distribution and some transistor design parameters (gate oxide thicknesses and dopant surface concentration). The theoretical analyses and computed illustrations are given in section 3.2. Analyses of the experimental data to give transistor design parameters in the MCR and $N_{IT}(y)$ profiles in both the JSCR and MCR are presented in sections 3.3, 3.4 and 3.5 respectively. A summary is given in section 3.6.

3.2 Theoretical Analysis

The analytical formulas were derived and described in chapter 3 for the Shockley-Read-Hall (SRH) steady-state recombination rate R_{SS} at interface traps in the MCR and JSCRs. There is an additional V_{GB}-dependence of R_{SS} from N_{IT} in the LDERs, which will be included in this chapter. The present analyses for the LDER follow the steps for the MCR with two noticeable differences: **(i)** the nearly zero flatband voltage in the LDERs due to the same dopant type and nearly the same high dopant impurity concentrations in the Si-gate and drain and source LDERs according to the traditional designs of digital CMOS transistors and **(ii)** the gate-to-drain voltage is

now $V_{GD} = V_{GB} - V_{DB} = V_{GB} \pm V_{PN}$ (+ for nMOST and – for pMOST) which gives I_B the additional V_{PN} dependency, both are leverageable for N_{IT} characterization described as follows.

As illustrated by the computed potential variation curves in Fig.1, the (V_{PN}) forward-bias-injected minority carriers reduce the interfacial potential barrier height on both the MCR-side (lower curves) and the LDER-side (upper curves) of the JSCR shown in Fig. 3.1(a), and the total potential barrier height through the JSCR shown in Fig. 3.1(b). This V_{PN} dependency significantly affects the DCIV or I_B-V_{GB} lineshape from N_{IT}'s located in the JSCR. [The interfacial electric potential, V(x,y) at the SiO2/Si interface x=0, has been denoted by the two traditional and two recent engineering notations: $\psi(x=0,y) \equiv \psi_S(y) \equiv V_I(x=0,y) \equiv V_{IS}(y)$. The subscript I for intrinsic Fermi potential will be dropped for arbitrariness of the reference electric potential and for notation compactness using kT/q normalization: $U_S(y) = (q/kT)V_{IS}(y)$.] The curves in Figs. 3.1(a) and (b) were computed using the carrier depletion as the zeroth approximation in the JSCR [45] but they do include, in the MCR and the LDERs, the injected carrier concentration from the forward bias (V_{PN}) just described, and also the Fermi-Dirac carrier distribution and impurity deionization for high impurity surface or interface concentrations [46]. The upper part of Fig. 3.1(a) gives the total lateral drop of the interfacial potential in the LDER with 10^{-19}cm ^{3}. The potential-drop is about 100 times smaller than that in the MCR with $10^{17}cm^{-3}$ shown in the lower part of Fig. 3.1(a) as expected from the concentration ratio. Figure 3.1(b) shows that at V_{GB}=−1V (about flatband in the MCR), the p/n junction barrier height, V_B, at the interface is reduced

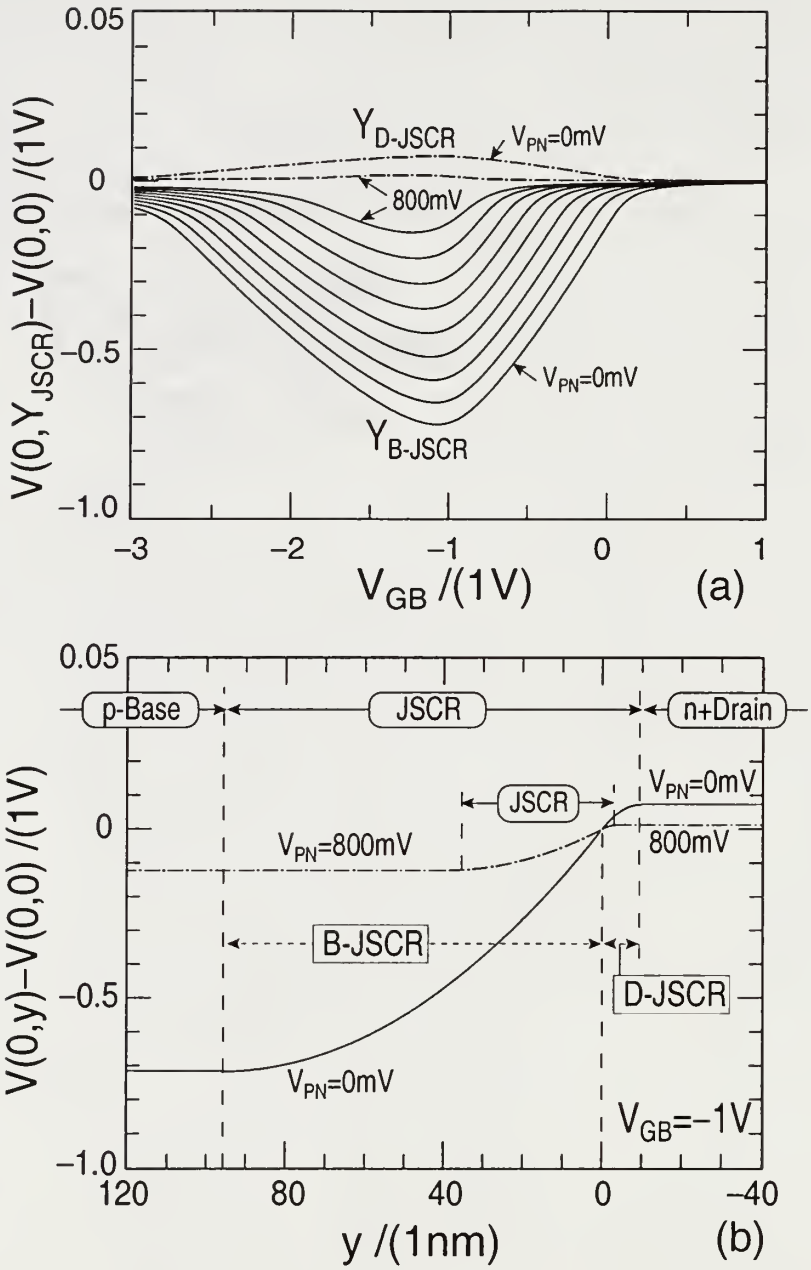

Fig. 3.1 Gate-Basewell voltage (V_{GB}) dependence of the potential variation through the space-charge layer at the SiO_2/Si interface of the drain or source n+/p-Base junction covered with a n+Si gate in the carrier-depletion and thick-oxide limits with spatially constant dopant impurity concentrations of $N_{GG}=5\times10^{19}cm^{-3}$, $N_{DD}=1\times10^{19}cm^{-3}$, and $P_{BB}=1\times10^{17}cm^{-3}$, and oxide thickness of $X_{OX}=3.0nm$. **(a)** The potential drop in the drain and basewell sides of the junction space-charge region, $V_{D\text{-}JSCR}$ and $V_{B\text{-}JSCR}$ for forward applied dc biases of $V_{PN}=0mV$ to 800mV. **(b)** The potential variation along the SiO_2/Si interface for $V_{PN}=0mV$ and 800mV.

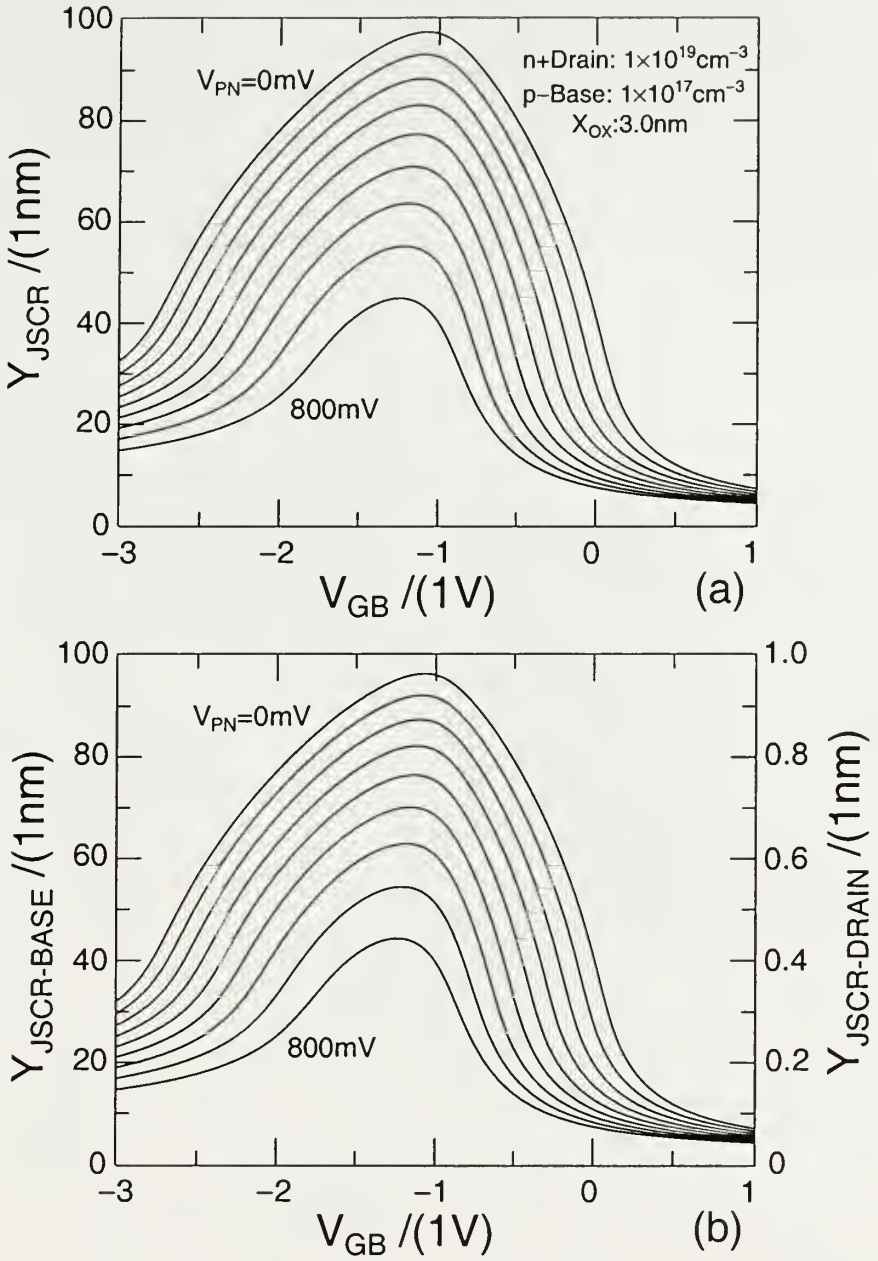

Fig. 3.2 The length of junction space-charge region as a function of the applied dc gate voltage, V_{GB}, for applied junction dc biases V_{PN}=0mV to 800mV. **(a)** Total length, Y_{JSCR}. **(b)** Length on the basewell and drain (or source) sides, $Y_{JSCR-Base}$ and $Y_{JSRC-Drain}$. Transistor parameters are the same as those used in Fig. 3.1.

nearly 600mV by the injected carriers, from V_B=720mV at V_{PN}=0mV to V_B=120mV at V_{PN}=800mV. The geometrical effect, shortening of the length of the JSCR by V_{PN} in Fig. 3.1(b), should also be noted which is leveraged in N_{IT} profiling. This is further illustrated in Fig. 3.2(a) for the total length, Y_{JSCR}, and in Fig. 3.2(b) respectively for Y_{JSCR}'s component on the Channel-side (left Y-axis) and the LDER- (or Drain- and Source-) sides (right Y-axis) whose curves are nearly the same with ratio 100 from the dopant impurity concentration ratio at the interface, $10^{19}/10^{17}$=100.

The effects of this lateral interfacial potential variation with forward bias are shown in Figs. 3.3(a) and 3.3(b) in which the normalized SRH recombination rate, R_{SS}/R_{SSpk}, or normalized basewell-pad current, I_B/I_{Bpk}, at a midgap interface trap in the JSCR are computed as a function of V_{GB}. As described in the following two paragraphs, the I_B-V_{GB} lineshape provides the basis for calculating the spatial dependence of the interface trap density, $N_{IT}(y)$ from experimental DCIV I_B-V_{GB} curves.

In Fig. 3.3(a), the effects of spatial variation of $N_{IT}(y)$ in the JSCR on the I_B-V_{GB} characteristics (at V_{PN}=200mV) are illustrated for the seven $N_{IT}(y)$ distributions: $N_{IT}(y)=N_{IT0}$=constant in the **(i)** MCR, **(ii)** Drain (or Source) LDER, and **(iii)** Drain (or Source) JSCR; and the four discrete $N_{IT}(y)=N_{IT1}\delta(y-Y_T)$ distributions located at **(iv)**-**(vii)** $y=Y_T$=10, 20, 30 and 40nm from the metallurgical p/n junction boundary. There are several obvious features in Fig. 3.3(a) useful for $N_{IT}(y)$ characterization from experimental data. First consider the six distributions except the constant-N_{IT} in the Drain- or Source-JSCR, (iii). **(1)** The V_{GBpk} moves from inversion to depletion-

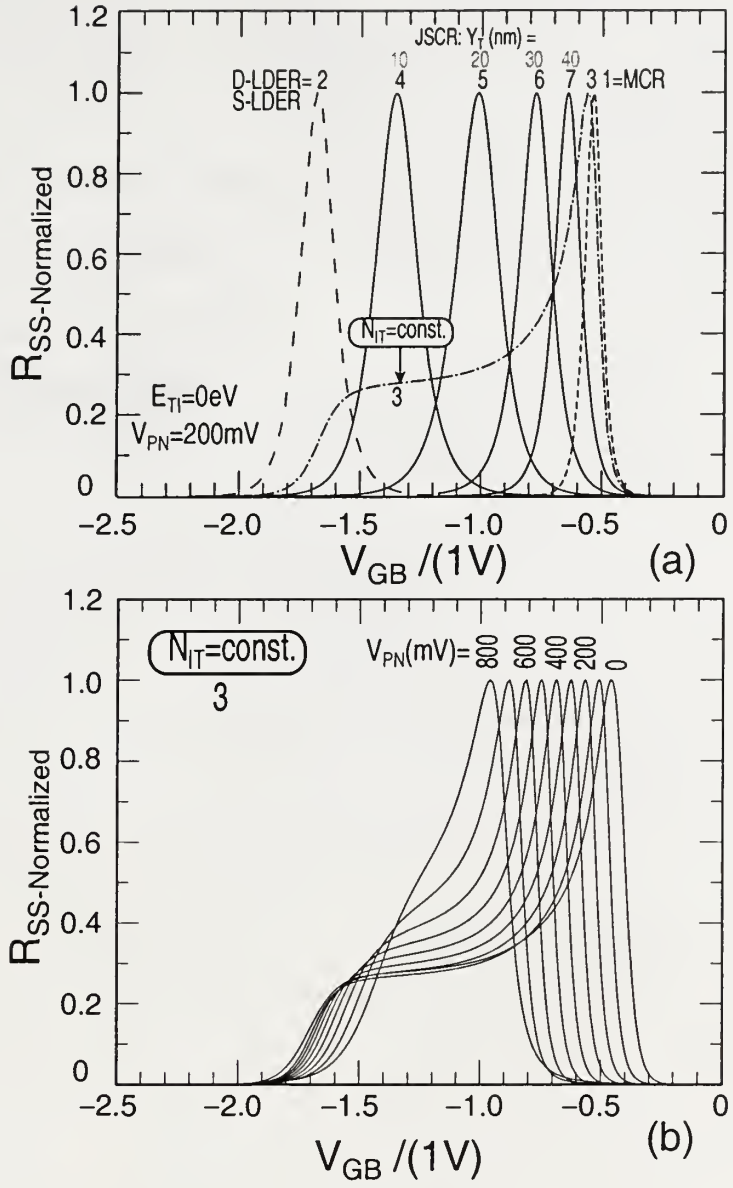

Fig.3 Effects of spatial variation of interface trap concentration (N_{IT}) and forward bias voltage (V_{PN}) on the normalized DCIV lineshape ($I_{B\text{-}Normalize}$ or $R_{SS\text{-}Normalize}$ vs V_{GB}) from an Si-midgap (E_{TI*}=0mV) interface trap located in the drain and source lowly-doped-extension regions (D-LDER and S-LDER), the junction space-charge-region (JSCR), and the mid-channel-region (MCR). (a) Three spatially constant N_{IT}'s: curve1=MCR, curve2=D-LDER and S-LDER, and curve3=JSCR, and four discrete (delta-function) N_{IT}'s located at Y_T=10, 20, 30, and 40nm (curve4,5,6,7) from the p/n impurity concentration boundary, at a forward-bias of V_{PN}=200mV. (b) Forward-bias dependence (V_{PN}=0mV to 800mV) for the constant N_{IT} in the JSCR.

flatband-accumulation as the discrete $N_{IT}(y=Y_T)$ moves from the MCR edge towards the p/n boundary y=0. **(2)** The I_B-V_{GB} lineshape from the constant-N_{IT} in the MCR and D-LDER (or S-LDER) and from the four spatially discrete $N_{IT}(y)$ all are sharply peaked Gaussian-like with increasing width as Y_T moves from the MCR towards the Drain (or Source) LDER. These two lineshape properties provide the basis for its utility as a delta-function probe for the lateral variation of $N_{IT}(y)$.

For the spatially constant $N_{IT}(y)=N_{IT0}$ in the JSCR, (iii), the following effects on the lineshape and position are noted. From the V_{PN}=200mV curve in Fig. 3.3(a): **(31)** the V_{GBpk} is near that of the N_{IT} located in the MCR (i), and **(32)** the upper part of the lineshape is similar to that of MCR-N_{IT} (i) also. The curves in Fig. 3.3(b) shows the effect of increasing V_{PN} or injected minority carrier concentration: **(33)** V_{GBpk} shifts toward flatband-accumulation (negative V_{GB} in nMOST), **(34)** the lineshape broadens, **(35)** a substantial plateau, about 25% for $V_{PN}\rightarrow 0$, diminishes, **(36)** the upper 20% of the I_B-V_{GB} lineshape is nearly Gaussian with a slightly skew on the accumulation side, and **(37)** the entire half of the I_B-V_{GB} lineshape on the inversion-side (positive V_{GB} in nMOST) is nearly Gaussian-like.

Figures 3.4 and 3.5 provide theoretical estimates of the confidence level or the accuracy of using the delta-function approach to analyze the DCIV I_B-V_{GB} data for extracting slowly and rapidly varying $N_{IT}(y)$ profiles respectively. The assumed (continuous lines) and the extracted (crosses) profiles are shown in Figs. 3.4(a) and 3.5(a), respectively for the exponential profile, exp(-y/3nm), and the Gaussian profiles, $\exp[-(y-Y_T)^2/L_T^2]$ located at four distances Y_T=10,20,30,40nm from the p/n boundary.

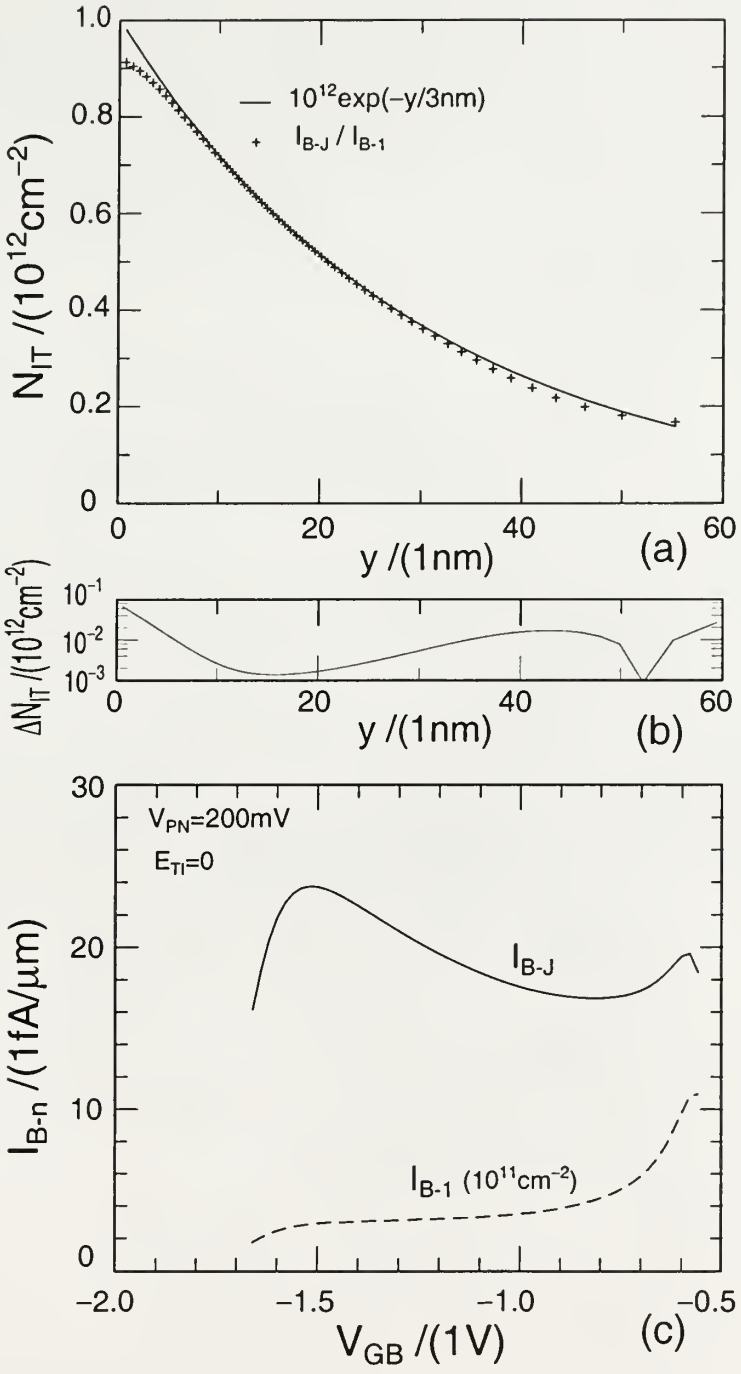

Fig. 3.4 Illustration of the use of the DCIV line-shape for profile extrapolation of the mid-gap interface trap concentration, $N_{IT}(y)$, for an exponential profile located inside the JSCR, at V_{PN}=200mV. Transistor parameters are the same as those of Fig. 3.1. **(a)** Assumed and extrapolated profiles. **(b)** Fractional error. **(c)** Current integrals of the profile-extrapolation algorithm.

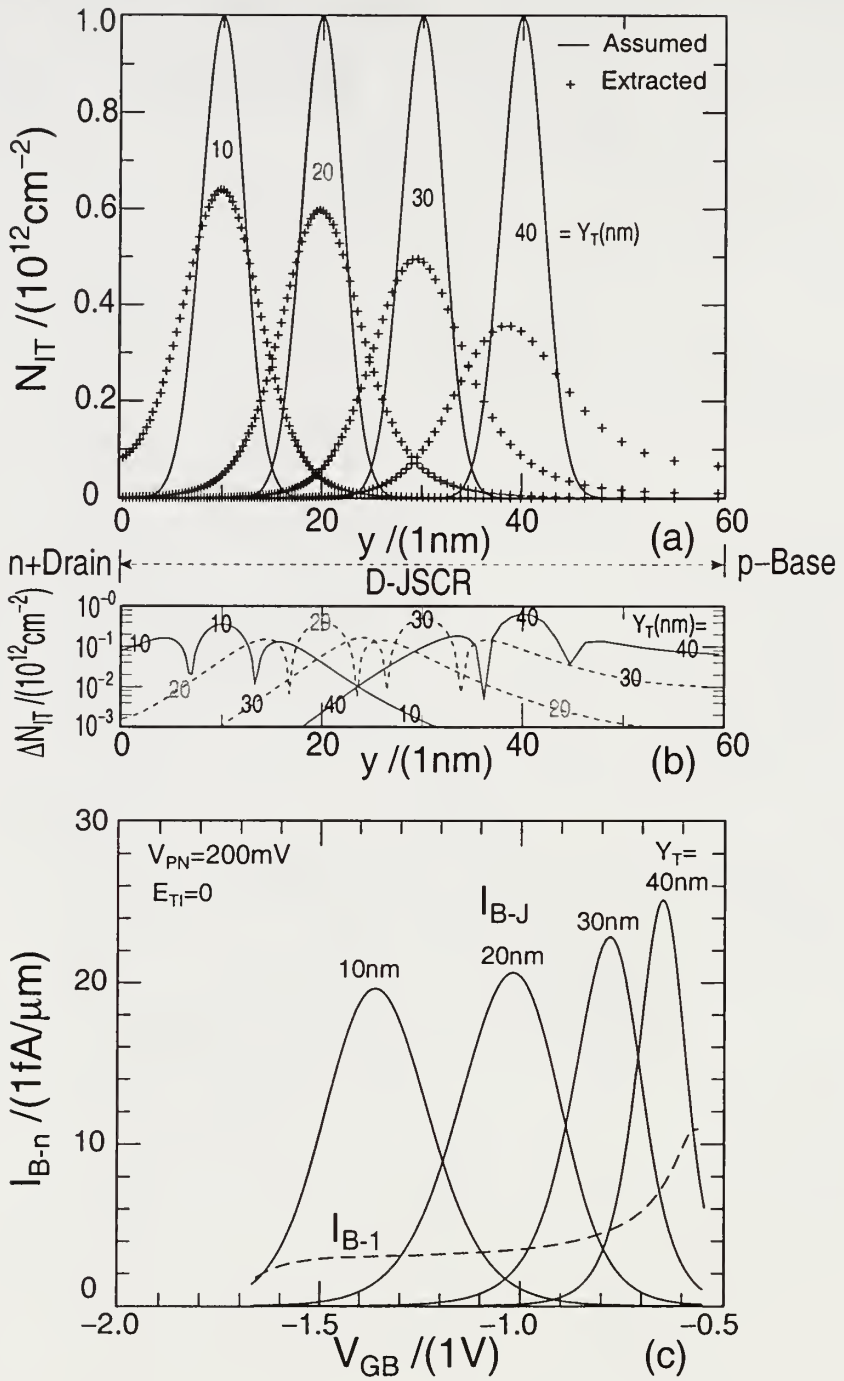

Fig. 3.5 Illustration of the use of the DCIV line-shape for profile extrapolation of the mid-gap interface trap concentration, $N_{IT}(y)$, for four Gaussian profiles located at Y_T=10, 20, 30 and 40nm from the n+D/p-B boundary inside of the JSCR at V_{PN}=200mV. Transistor parameters are the same as those of Fig. 3.1. **(a)** Assumed and extrapolated profiles. **(b)** Fractional errors. **(c)** Current integrals of the profile-extrapolation algorithm.

The difference or fractional error of the extracted profiles are shown in Figs. 3.4(b) and 3.5(b), indicating excellent accuracy (<0.07 or 7%) for the slowly varying exponential N_{IT} in Fig. 3.4(b), but rather large differences for the rapidly varying Gaussian N_{IT} in Fig. 3.5(b), the latter as expected since $R_{SS1}(y,V_{GB})$ is far from a delta-function, nevertheless, Fig. 3.5(a) shows that extrapolated profile are peaked and Gaussian-like with a peak location near the true physical location and with a linewidth about 1.5-times wider which increases ~2-times wider as the peak moves towards the base-side or channel-side of the JSCR. Thus, theoretical corrections could be made to correct part of such a systematic difference.

The $N_{IT}(y)$ profile extraction procedure, used to give Figs. 3.4(a), 3.4(b), 3.5(a) and 3.5(b), starts with the delta-approximation for evaluating the general formula (2.21)

$$I_{B-JSCR}(V_{GB}) = qW\int_{0}^{Y_{JSCR}} N_{IT}(y)\cdot R_{SS1}(y,V_{GB})\,dy \qquad (3.1)$$

where W is the channel width. The integrand $R_{SS1}(y,V_{GB})$ is the SRH recombination rate per unit interface trap. Its lineshape in y and V_{GB} are both bell-shaped and rather sharp, resulting in a simple but accurate mapping of the experimental I_B-V_{GB} to the $N_{IT}(y)$-y profile. The bellshape characteristics are illustrated by the following examples. The maximum of $R_{SS1}(y)$-y at a given V_{GB} [See Fig. 3.6(a).] occurs at the position $y=Y_M(V_{GB})$ which was given by Eq. (2.4):

$$U_{S*J} = U(Y_M)-(U_p+U_N)/2 + \log_e(c_{ns}/c_{ps})^{1/2} = 0 \qquad (3.2)$$

U_P and U_N are the hole and electron quasi-Fermi potentials, respectively. Using the U(y) from the depletion approximation and the abrupt dopant impurity concentration

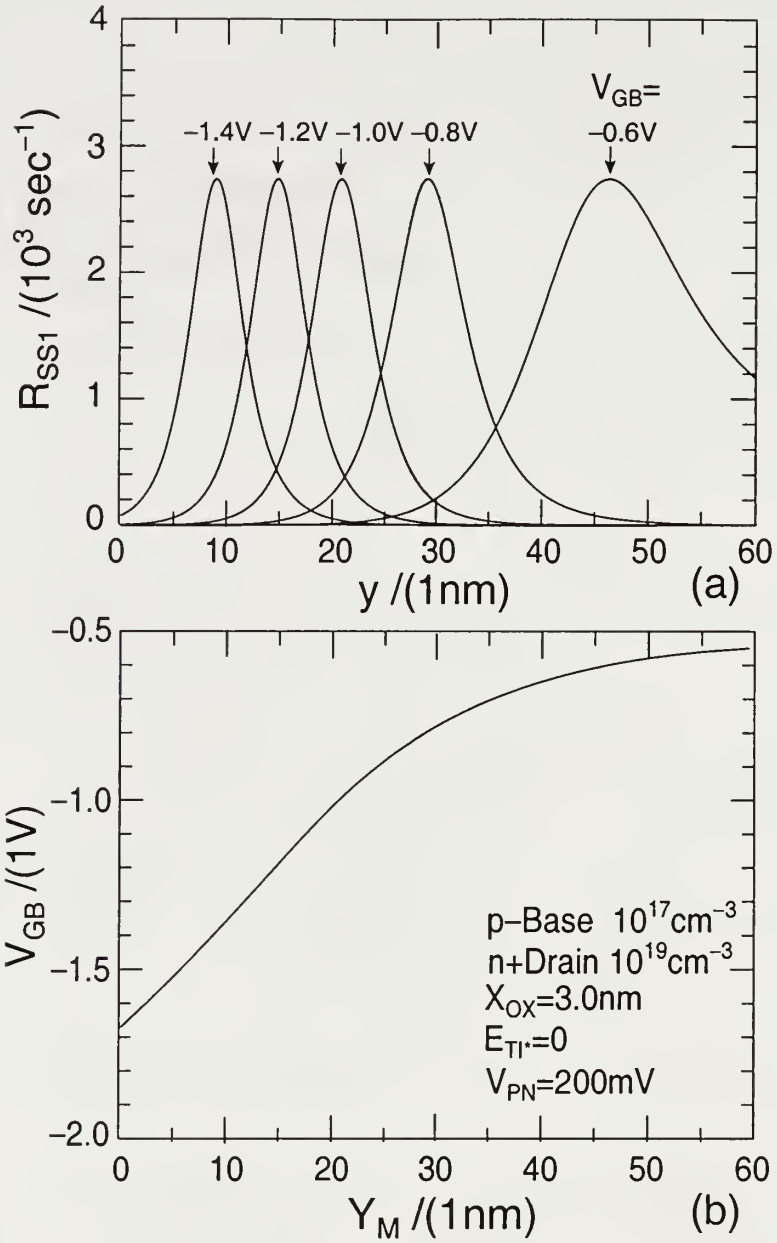

Fig. 3.6 Gate-voltage dependence of the spatial variation of the steady-state recombination rate at V_{PN}=200mV. Other parameters are the same as those of Fig. 3.1. **(a)** Gate voltage dependence of R_{SS1} vs y. **(b)** Location, Y_M, of the peak R_{SS1} vs gate voltage.

model, Y_M has the following dependence on V_{GB} where U_{SX} and U_{SE} are total surface potential changes in the x-direction normal to the SiO_2/Si interface in the MCR (Basewell) and the D-LDER or S-LDER (Emitters) respectively.

$$Y_M(V_{GB}) = \sqrt{2}\, L_D \left[\sqrt{U_{BI}-U_{SX}-U_{PN}+U_{SE}} - \sqrt{U_F-U_{PN}/2-U_{SX}-\log_e(c_{ns}/c_{ps})}^{1/2}\right] \quad (3.3)$$

L_D is the Debye length, U_{BI} the built-in potential of the n+D/p-B junction. This equation maps a gate voltage to a spatial location in the JSCR in order to extract the $N_{IT}(y)$ profile from the experimental I_B-V_{GB}. Shown in Fig. 3.6(a) for the nMOST with midgap trap at V_{PN}=200mV, as V_{GB} is swept from subthreshold-inversion (V_{GB}=–0.6V) to flatband-accumulation (V_{GB}=–1.4V), the position of the R_{SS1} maximum, Y_M, moves from the MCR-edge of the JSCR (Y_M=46nm) towards the metallurgical p/n boundary or the LDER-edge of the JSCR (Y_M=9nm $\rightarrow$ 0). This shift is computed from (3.3) and shown in Fig. 3.6(b) indicating almost a linear dependence of 30nm/V in the flatband-accumulation range (V_{GB}<–0.94V). Another feature is the narrowing of the $R_{11S}(y)$ linewidth towards flatband-accumulation shown in Fig. 3.6(b) which sharpens the probe for the $N_{IT}(y)$ profile. The sharp bellshaped $R_{SS1}(y)$ can be approximated by a delta-function to evaluate the I_B integral, (3.1):

$$I_{B\text{-}JSCR}(V_{GB}) \approx qWN_{IT}(Y_M)\int_0^{Y_{JSCR}} R_{SS1}(y,V_{GB})\,dy \equiv N_{IT}(Y_M)*I_{B\text{-}1}(V_{GB}) \quad (3.4)$$

where $I_{B\text{-}JSCR}$ is the experimental base-terminal current or theoretically generated current using (3.1). The **interface trap profiling formula in the JSCR** is then given by

$$N_{IT}(Y_M) = I_{B\text{-}JSCR}(V_{GB})/I_{B\text{-}1}(V_{GB}) \quad (3.5)$$

where $I_{B-1}(V_{GB})$ defined by (3.4) is the recombination current from a spatially constant interface trap concentration inside the JSCR normalized to $N_{IT}(0<y<Y_M)=1$.

To estimate the error in the extracted N_{IT} profile using the delta-function approximation in (3.4), an assumed profile of N_{IT} in the JSCR is used to compute $I_{B-J}(V_{GB})$ from (3.1) and $I_{B-1}(V_{GB})$ from (3.4). These were given in Fig. 3.4(c) for the exponential profile and Fig. 3.5(c) for the four Gaussian profiles. The ratio of these two curves then give the extracted N_{IT} profiles in Figs. 3.4(a) and 3.5(a) and the corresponding difference errors in Figs. 3.4(b) and 3.5(b). Practical applications using this procedure to extract pre-stress and post-stress N_{IT} profiles are given in section 3.4.

Figures 3.7-3.9 give additional illustrations of the dependencies of the DCIV I_B-V_{GB} characteristics on the forward-bias V_{PN} or injected minority carrier concentration. These V_{PN} dependencies can be used to determine additional properties of the interface traps. The trap-spatial-location effects on the V_{GBpk}-V_{PN} is shown in Fig. 3.7 for the N_{IT} distributions given in Fig. 3.3(a) which indicate that increasing forward bias, V_{PN}, or minority carrier concentration would decrease V_{GBpk} towards flatband-accumulation if $N_{IT}(y)$ is constant in the JSCR and MCR, or localized near MCR (Y_T>30nm) while increase V_{GBpk} towards subthreshold and inversion if $N_{IT}(y)$ is localized (delta-function) in the JSCR (Y_T<30nm) or in the LDER.

The trap-energy-location effects on the shape of the R_{SS} or I_B vs V_{GB} curve is shown in Fig. 3.8(a) for a spatially constant N_{IT} in the JSCR and four trap energy levels, E_{TI}=0.0V (midgap), 0.1V, 0.2V and 0.3V. The shallower traps (E_{TI}=0.2V, 0.3V) have a broader R_{SS}-V_{GB} or I_B-V_{GB} and steeper and higher plateau in the flatband-accumulation

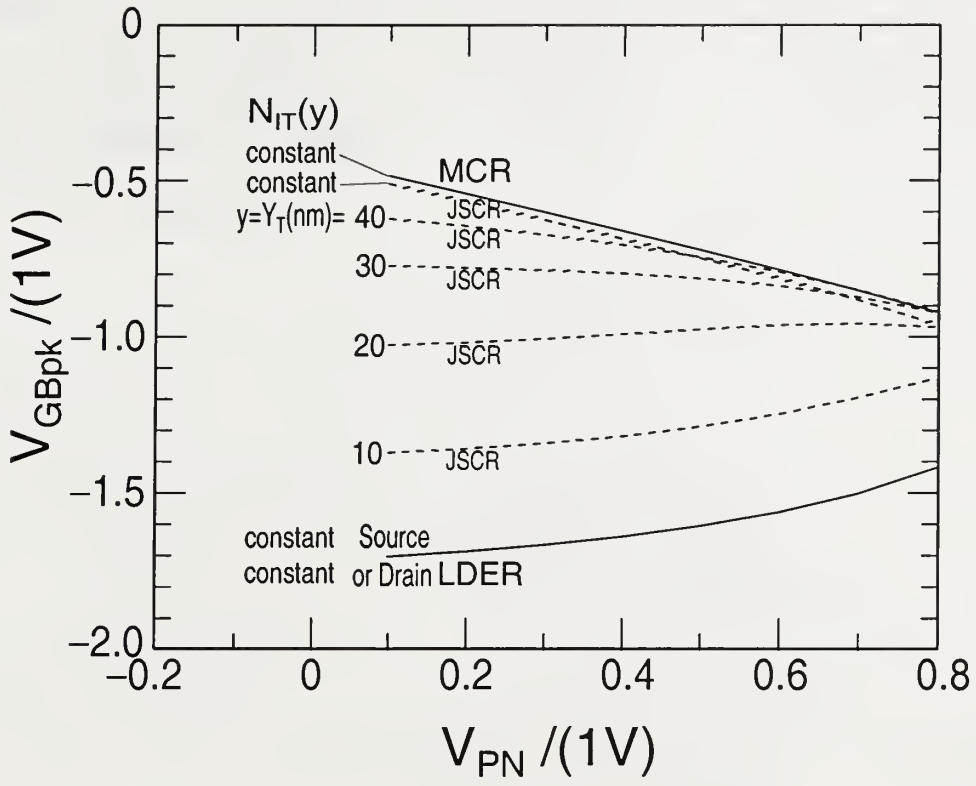

Fig. 3.7 Forward-bias (V_{PN}) dependence of the V_{GB} at I_{Bpk} for the seven $N_{IT}(y)$ distributions shown in Fig.3.

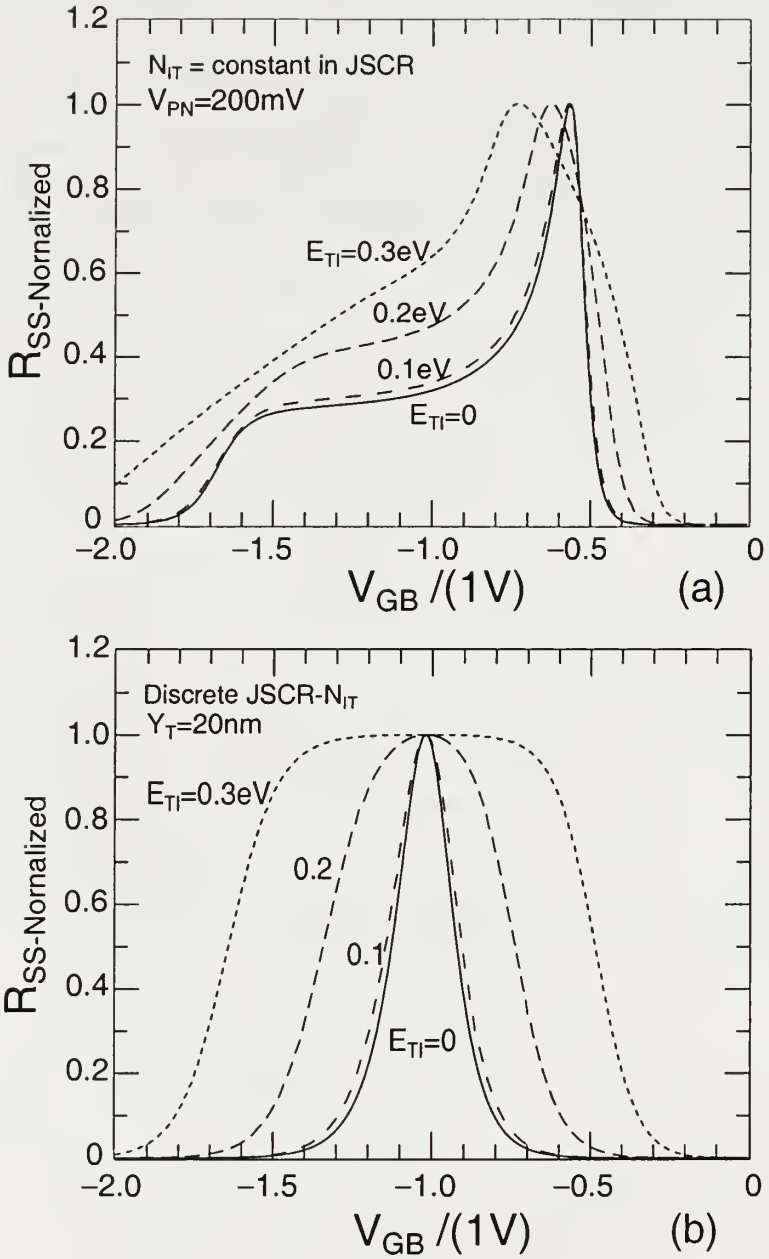

Fig. 3.8 Effects of energy level of the interface trap in the JSCR on the DCIV lineshape at V_{PN}=200mV. Other transistor parameters are the same as those in Fig. 3.1. **(a)** Constant N_{IT} in the JSCR. **(b)** Discrete N_{IT} at Y_T=20nm from the n+D/p-B boundary.

range (V_{GB}<−1V in this nMOST example). Figure 3.8(b) is for localized N_{IT} in the JSCR which shows broadened I_B-V_{GB} when the energy level of the interface trap moves from the midgap E_{IT} = 0 towards one of the band edge, E_{IT} = 0.3eV.

The trap-energy-location effects on the shape and the normalized reciprocal slope of the log(I_{Bpk}) vs V_{PN} characteristics are shown in Figs. 3.9(a) and 3.9(b). (We called this the Peak Plot or PP in contrast to the Gummel Plot and the Shockley Plots.) This PP is probably the most powerful result of the DCIV methodology for rapid interface trap energy level determination because the V_{PN}-dependence of the peak I_B current is a function of only the effective interface trap energy level, $E_{TI*} = E_{TI} + (kT)\log_e(c_{ns}/c_{ps})^{1/2}$. It is completely independent of the surface energy band bending and other material properties, such as surface impurity concentrations, oxide thickness, and carrier capture rates. Although E_{TI*} contains the carrier capture rate ratio it would only give an uncertainty in the extracted E_{TI} of $(kT/q)\log_e(c_{ns}/c_{ps})^{1/2} \sim 1.1(kT/q) \sim$ 30mV since the carrier capture ratio is usually in the range of 0.1 to 10 for neutral trap potential wells of the intrinsic and silicon-dangling-bond interface traps observed experimentally thus far. Figure 3.9(a) gives the I_{Bpk}-V_{PN} plots for four E_{TI}(eV) (0, 0.1, 0.2, 0.3) with the seven spatial N_{IT} distributions of Fig. 3.3(a). Figure 3.9(b) gives the reciprocal slope defined by

$$\exp(U_{PN}/n) - 1 = [\exp(U_{PN}) - 1] / [\exp(U_{PN}/2) + \cosh(U_{TI*})] \qquad (3.6)$$

Six of the N_{IT} distributions have identical PP's (unbroken lines) in Fig. 3.9(a) which are the two constant-N_{IT}'s in the MCR and LDERs (Drain and Source), and the four delta-function-N_{IT}'s in the JSCR. Their n-factor, shown Fig. 3.9(b), approach the ideal

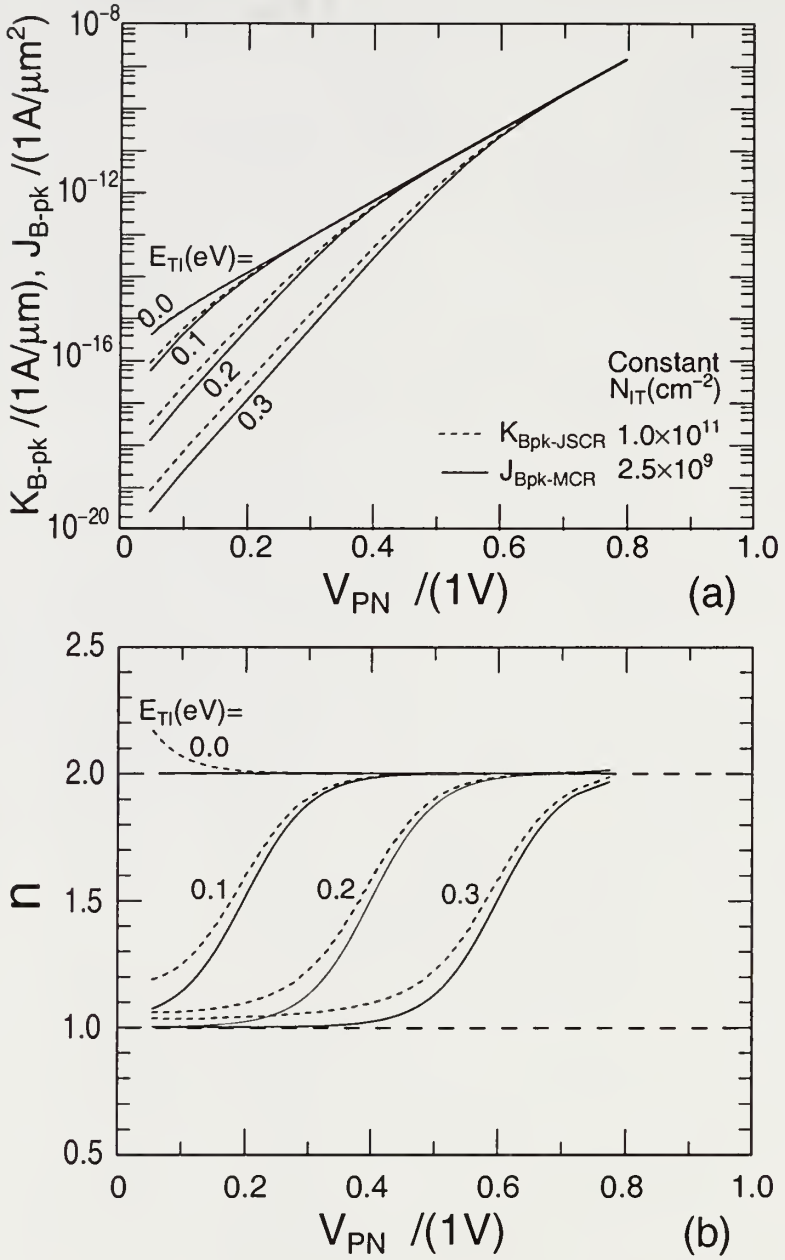

Fig. 3.9 Effects of energy level of the interface trap in the JSCR and MCR on the peak recombination current vs forward bias. Other transistor parameters are the same as those in Fig. 3.1. **(a)**I_{Bpk} and K_{Bpk} vs V_{PN}. **(b)** Reciprocal slope or voltage-swing vs V_{PN}.

Shockley value of n=1 at low V_{PN} for the three non-midgap traps (E_{TI*} = 0.1eV, 0.2eV, 0.3eV), increasing to n=2 at high V_{PN} or $P_S=N_S >> n_i$ in a small range of V_{PN} about 200mV (10% to 90% rise). It degenerates to n=2.0 at all V_{PN} for the midgap trap $E_{TI*}=0$. The seventh, constant-N_{IT} in the JSCR, has a slightly higher PP (broken line) in the lower V_{PN} range and a higher n than the other six distributions. For this case, Fig. 3.9(b) also illustrates the contribution to I_{Bpk} from the N_{IT} at the channel or MCR-edge of the JSCR. This geometric effect from $Y_{JSCR}(V_{PN})$ increases n above 2 at low V_{PN} for the midgap trap as indicated by the dotted line in Fig. 3.9(b). To remove this geometric effect, the $Y_{JSCR}(V_{PN})$ dependence can be taken out approximately by normalizing the numerically integrated I_{Bpk} integral to $Y_{JSCR}(V_{PN})$ before using (3.6) to compute n. Such a procedure is desired in this theoretical illustration since the reciprocal slope is defined for the purpose of quickly assess the properties of the interface traps from the experimental V_{PN} dependence of I_{Bpk} used in the initial analyses of the gate-controlled base-terminal current [5,8] and the 1957 SNS theory for the recombination at traps in the bulk JSCR [40].

The insensitivity on the spatial distribution of N_{IT} of the six distributions first considered above allows an accurate determination of the effective energy level of the interface traps, E_{TI*}, and the thermal activation energy of the I_B which is an important component of the standby current in CMOS operations. In practice, the interface traps have always been observed near the midgap ($E_{TI*}\simeq 0$eV). Some extracted shallow-trap energy levels in previous analyses of data [8,33,34] originated from emitter-injection-

limited and diffusion-limited I_B current through a thick basewell layer or long channels rather than from shallow interface traps.

3.3 Extraction of MOST Device Parameters

One of the applications of the DCIV base-recombination-current methodology was its use as the diagnosis monitor for transistor design and fabrication processes [8]. In this section, a demonstration is given on the extraction of the surface concentration of the base dopant impurities, N_{XX}, and the oxide thickness, X_{OX}, and their comparisons with the design values. A sample of the experimental Top-Emitter DCIV (TE-DCIV) data matrix used in this demonstration is shown in Figs. 3.10(a) and (b) for two drawn channel lengths, W/L=10μm/10μm and 10μm/0.25μm, of pre-stressed pMOSTs on 8-inch wafers fabricated by a 0.25μm technology. The other 4-sets in the 6-set matrix were obtained for L = 5, 2, 0.8 and 0.4μm drawn channel lengths. Similar application results have been obtained for earlier (0.5μm and 0.35μm) and later (≤0.18μm) technologies [8].

The broke lines in Figs. 3.10(a) and (b) show the location of the (I_{Bpk},V_{GBpk}) whose shift towards flatband-accumulation (positive V_{GB} for pMOSTs) with increasing drain and source forward-bias, V_{PN}=0-700mV, is characteristic of an interface traps in the MCR. The I_B-V_{GB} curve around the I_{Bpk} is least-squares-fitted (LSF) to a Gaussian with a baseline,

$$I_B = A_1 + A_2 \cdot \exp\{-[(V_{GB}-A_3)/A_4]^2\} \qquad (3.7)$$

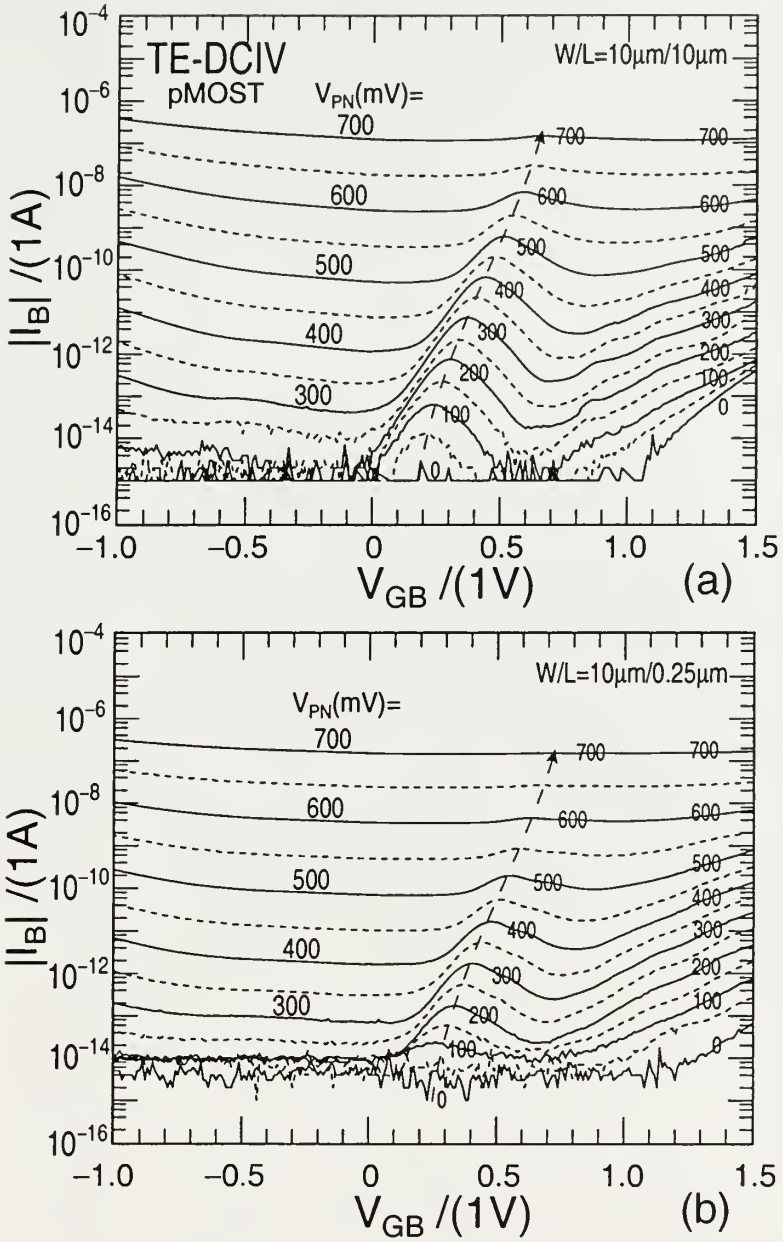

Fig. 3.10 Experimental TE-DCIV of pMOSTs from quarter-micron CMOS technology for V_{PN}=50mV to 700mV. Channel Width to Length aspect ratio W/L of **(a)** 10μm/10μm. **(b)** 10μm/0.25μm.

The four fitting parameters are the baseline $A_1=I_{B0}$, the peak current $A_2=I_{Bpk}$, the peak gate voltage $A_3=V_{GBpk}$ and the linewidth A_4. These four LSF parameters were obtained for the 84 DCIV curves from pMOSTs with the six L at the fourteen V_{PN} values (50, 100, ..., 650, 700mV). Figure 3.11 shows a typical LSF result for the L=10μm and 0.25μm pMOST at V_{PN}=400mV. The simple Gaussian cannot be used to fit the entire gate voltage range because (**i**) the DCIV lineshape from N_{IT} in MCR is skewed toward the flatband-accumulation side of the peak (towards positive V_{GB} in pMOSTs), with a slightly wider half-width-half-maximum (HWHM) as previously shown chapter 3 and the skew was also observed in the analytical solution shown in Fig. 3.3(b) for a constant N_{IT} in the JSCR and (**ii**) constant or peaked N_{IT} in the JSCR contributes to additional I_B to that from N_{IT} in the MCR thereby lineshape widening on the flatband-accumulation side of the MCR peak. To circumvent these two asymmetry possibilities, only data points from the inversion range and a small range on the accumulation side of the I_{Bpk} are LS Fitted to (3.6). These are the open circles (L=10μm) and squares (L=0.25μm) shown Fig. 3.11. The difference, $I_{B\text{-}DATA} - I_{B\text{-}FIT}$, are shown by the two lower-height humps labeled JSCR in Fig. 3.11 which are from the N_{IT} distributions in the drain and source JSCRs or LDERs. The V_{GBpk} (or A_3) from the LSFs are plotted as a function of V_{PN} in Figs. 3.12(a) and 3.12(b) for the main I_{Bpk1}. The increasing V_{GBpk1} towards flatband-accumulation with increasing V_{PN} (pMOST) suggest that the I_{Bpk1} arises from N_{IT} located in the MCR. Thus, we proceeded to fit the (V_{GBpk1},V_{PN}) datapoints to the analytical formulas developed in chapter 3:

$$V_{GBpk} = V_{FB} + V_{Spk} + 2 \cdot \mathrm{sign}(V_{Spk}) \cdot [V_{XX}(|V_{Spk}|-kT/q)]^{1/2} \qquad (3.8)$$

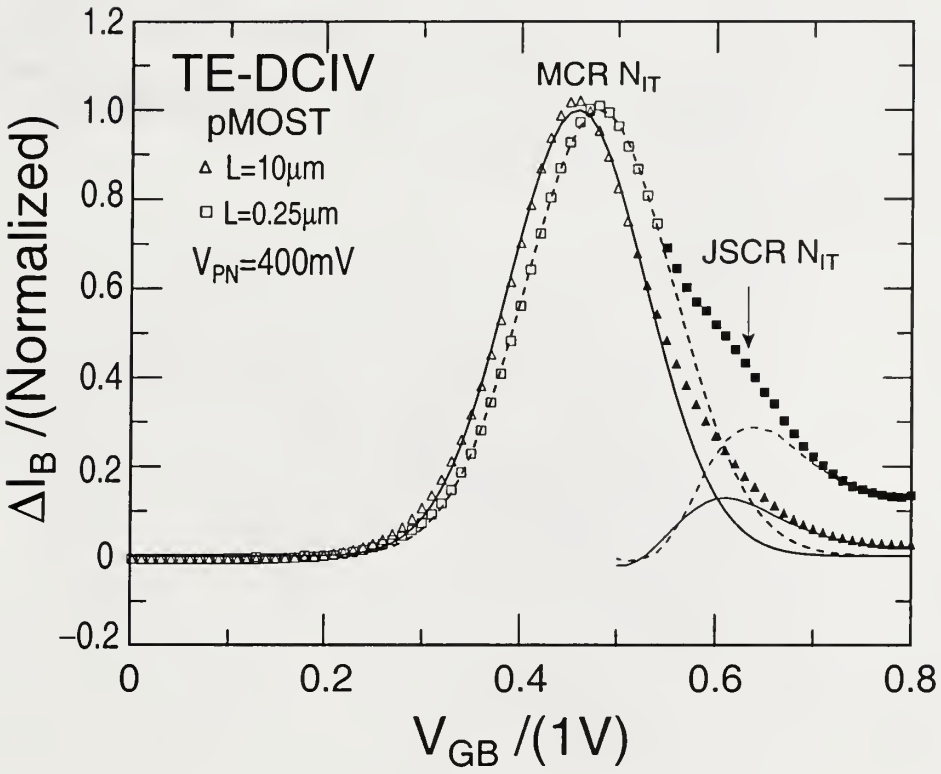

Fig. 3.11 Least-squares-fits of the TE-DCIV curve of L=10μm and 0.25μm pMOST at V_{PN}=400mV shown in Fig. 3.10. Decomposition into two components: one from N_{IT} in the MCR and one from N_{IT} in the JSCR.

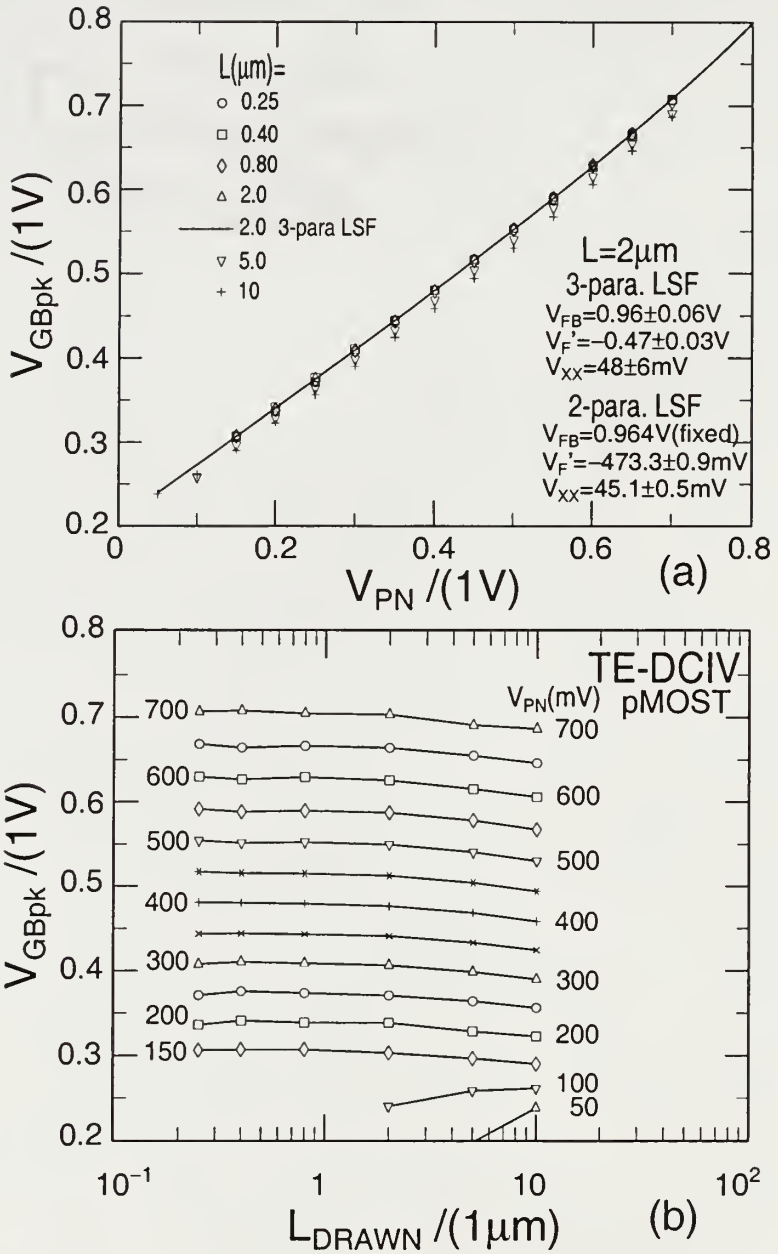

Fig. 3.12 Experimental channel length dependence of V_{GBpk} at I_{Bpk} at six L(μm) from 0.25 to 10 for extrapolation of dopant impurity concentration at the oxide/silicon interface. **(a)** V_{GBpk} vs V_{PN} with 3-parameter and 2-parameter least-squares-fit for L=2.0μm. **(b)** V_{GBpk} vs L for V_{PN}=50mV to 700mV.

where

$$V_{Spk} = V_F \pm V_{PN}/2 - \log_e(c_{ns}/c_{ps})^{1/2} \qquad (3.9)$$

where V_{Spk} is the Si substrate surface energy band bending at the peak, $V_{XX} = \varepsilon_S q N_{XX}/2C_{OX}^2$ where $C_{OX} = \varepsilon_{OX}/X_{OX}$, and $V_F = (kT/q)\cdot\log_e(N_{XX}/n_i)$ is the majority carrier quasi-Fermi potential. The term (-kT/q) in (8) was dropped in (2.11) and is retained here to provide better accuracy at high V_{PN}'s where $V_{S\text{-}pk}$ approaches zero or the flatband condition as indicated by the decreased drain/base p/n junction barrier height at the interface shown in Fig. 3.1. Combining (3.8) and (3.9) gives the 3-parameter formula to fit the experimental V_{GBpk}-V_{PN} data:

$$V_{GBpk} = A_1+A_2+(V_{PN}/2)-2\cdot\{A_3\cdot[-A_2-(V_{PN}/2)-(kT/q)]\}^{1/2} \qquad \text{(pMOST)} \quad (3.10)$$

The three LSF parameters are: $A_1 = V_{FB}$, $A_2 \equiv V_F' = V_F - (kT/q)\cdot\log_e(c_{ns}/c_{ps})^{1/2}$ and $A_3 = V_{XX}$.

As an example of the LSF procedure we discuss the steps to fit the L=2μm data to (10) which gives the solid line in Fig. 3.12(a). Only 9 (V_{PN}=150mV-550mV) of the 14 datapoints (50-700mV) were used which are the unfilled triangles of the L=2μm data in Fig. 3.12(a), in order to limit the experimental V_{GBpk} uncertainty to <0.5mV. For $V_{PN} \leq 100$mV, V_{GBpk} errors are larger than 0.5mV due to the ~3fA noise in I_B. In the $V_{PN} \geq 600$mV range, a more tedious fit to parametric equations instead of (3.10) is needed due to the high injection level. The 9-point V_{GBpk} data fits the 3-parameter equation (3.10) very well, as shown visually in Fig. 3.12(a), with a Chi-Square of only 0.16%. The LSF value of V_{FB}, V_F' and V_{XX} are also listed in Fig. 3.12(a). The

V_{FB}=0.96V is consistent with the designed polysilicon p+Gate for pMOST. The electron/hole capture rate ratio c_{ns}/c_{ps} is on the order of unity for near mid-gap traps, therefore $A_2 = V_F' \simeq V_F$, which gives a surface concentration for the n-Basewell donor (implanted As or P) impurity of $N_{XX}=N_{DD}=1.1\times10^{18}cm^{-3}$ which is consistent with the nominal design value of $\sim10^{18}cm^{-3}$. The gate oxide thickness can then be determined from $A_3=V_{XX}$ which gives X_{OX}=2.4nm compared with the nominal design value of 2.5nm. The exponential dependence of N_{XX} on the Fermi potential V_F gives a much larger uncertainty in X_{OX} and N_{XX} from the 6% uncertainty in V_F. The error propagates as follows:

$$\Delta N_{DD}/N_{DD} = \Delta V_F/(kT/q) \qquad (3.11)$$

$$\Delta X_{OX}/X_{OX} = (\Delta V_{XX}/V_{XX} + \Delta N_{DD}/N_{DD})/2 \qquad (3.12)$$

The data was taken at room temperature giving kT/q~25mV and using ΔV_F=0.03V, the percentage error in N_{XX} is 116% and X_{OX}, 64% in this global 3-parameter fit. However, if one of the three parameters (V_{FB}, N_{XX} and X_{OX}) is known from another independent measurement or is fixed in a regional 2-parameter fit, then the two LSF parameters may be determined more precisely. For example, we fixed the value of V_{FB} at 0.964V which was obtained by the 3-parameter fit, and then we made a 2-parameter LSF to (3.10). The results are listed in Fig. 3.12(a) showing a large reduction of uncertainties in V_F' and V_{XX}. The percentage errors translated to N_{XX} and X_{OX} are only 4% and 2% respectively.

To test the sensitivity of this 2-parameter extraction to the fixed value of V_{FB}, we vary V_{FB} by ±2% (~±20mV). An alternative test would be to fix X_{ox} from an

independent measurement, such as the C-V. The value of the device parameters (N_{XX} and X_{OX}) and their uncertainties, computed from the A's from the 2-parameter LSF with V_{FB} fixed, are summarized in the following table:

V_{FB}(V)	N_{XX}($10^{18}cm^{-3}$)	X_{OX}(nm)	χ^2(%)
0.945	0.76±0.02	2.87±0.06	0.15
0.964	1.13±0.04	2.40±0.05	0.15
0.984	1.68±0.06	2.00±0.05	0.15

The above table shows that the DCIV method can extract N_{XX} and X_{OX} to an accuracy of <10% provided that the flatband voltage is known to better than about (kT/q)/5 ~ 5mV. In the preceding analysis, assumption is made that $c_{ns}=c_{ps}$ so $V_F'=V_F$. In factory application, a calibration transistor with known substrate doping concentration and oxide thickness may be used to determine the c_{ns}/c_{ps} ratio, from which V_F can be extracted and the gate dopant impurity concentration can be computed from $V_{FG}=V_{FB}+V_F$ where V_{FG} denotes the Fermi-potential in the p+Gate. For $|V_{FG}| \geq 0.50V$, accurate gate doping concentration can be obtained only by including impurity deionization and Fermi-Dirac statistics. A similar fit to the I_{Bpk2} can be made to give the P_{XX} and X_{OX} in the LDERs.

3.4 Profiling Interface Traps in Junction Space-Charge Regions

Examples are now given on profiling the interface traps in the drain and source junction space-charge regions before and after channel-hot-carrier (CHC) stress. An earlier profiling technique employed the transient charge-pump current using its dependence on the drain reverse biased [47] which was shown to correlate with the DCIV method [48]. A recent demonstration of the DCIV technique for profiling

interface traps near the drain junction measured the differential basewell-terminal conductance by varying the drain reverse bias [43]. The differential DCIV method had a higher sensitivity and resolution and could profile a large portion of the channel. However, these methods are all limited to the region outside of the reverse-biased JSCRs. The new approach, as described in section 3.2, gives the spatial N_{IT} variation inside the forward-biased drain or source JSCR directly from the ratio of the experimental DCIV curve and the reference theoretical DCIV curve corresponding to a constant N_{IT} distribution. The rather good spatial resolution and accuracy of this new approach were already illustrated by using analytical $N_{IT}(y)$ models in Figs. 3.4-3.5 and described in section 3.2. Oxide charge built-up, which is usually associated with interface trap generation during CHC stress under large drain reverse-bias [32,49], changes the local flatband voltage which complicates data analysis. In this demonstration, the CHE stress is kept at low drain voltages to prevent injection of secondary hot holes over the SiO_2/Si barrier (4.25eV for hot holes and 3.12eV for hot electrons) [32].

The example to be given is the extraction of $N_{IT}(y)$ in a 0.35μm-nMOST testbed for reliability assessment of low-K interconnect materials [6]. The specific nMOST in this demonstration has a drawn aspect ratio of W/L = 50μm/0.4μm. The pre-stress and post-stress DCIV data are plotted in Fig. 3.13(a) with the V_{GB}-independent $I_{B\text{-baseline}}$ subtracted. The sharp pre-stress peak, I_{Bpk}≈0.6pA, at V_{GBpk}=−0.5V is in the subthreshold range and comes from recombination at the interface traps in the MCR. This peak gate voltage is consistent with the nominal X_{OX}=7.0nm and

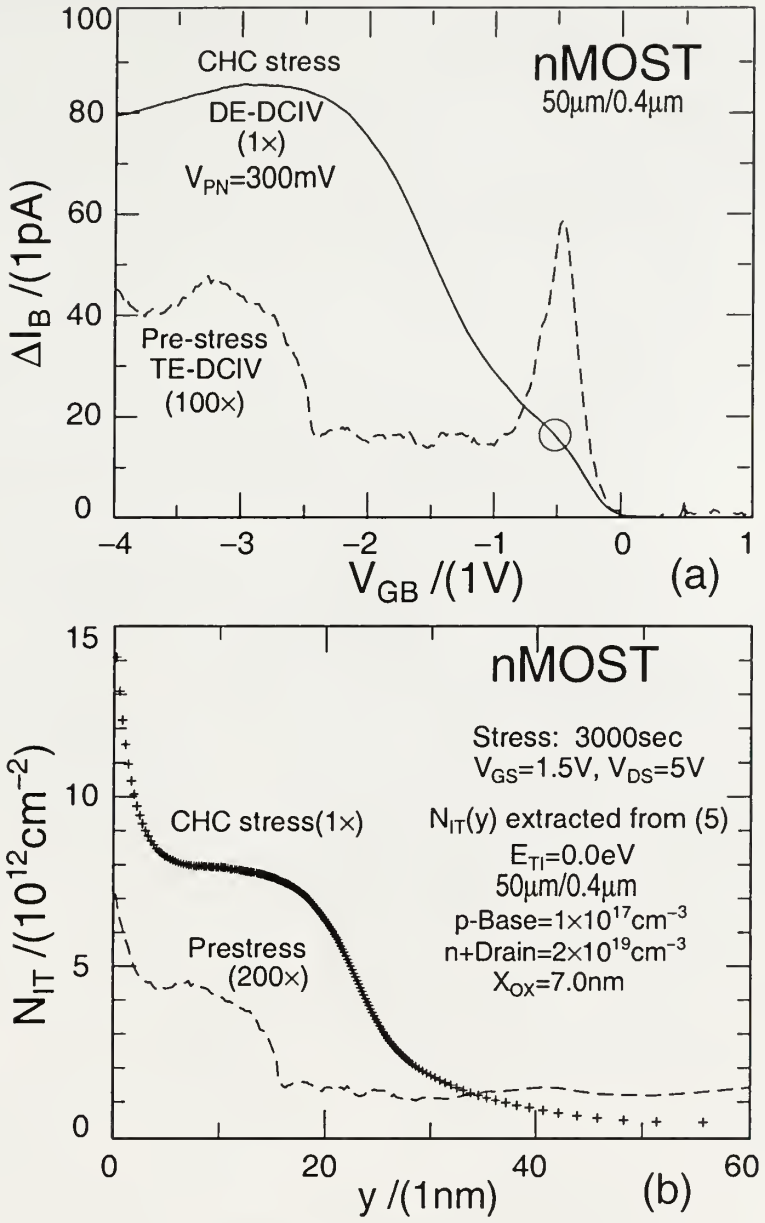

Fig. 3.13 Demonstration for the extrapolation algorithm applied to channel-hot-electron stress-generated interface-trap concentration profile of an nMOST from a 0.35μm low-k-interconnect CMOS technology. **(a)** Prestress TE-DCIV and poststress DE-DCIV curves. **(b)** Prestress and poststress $N_{IT}(y)$ profile extrapolations.

P_{XX}(boron)=$10^{17}cm^{-3}$. The broad pre-stress structure (~0.45pA) shown in Fig. 3.13(a) in the range V_{GB}<−2.5V (flatband-accumulation range for this nMOST) is from recombination at N_{IT} in the JSCR close to the metallurgical n+Drain/p-Basewell boundary y=0 at the SiO_2/Si interface x=0. This nMOST was then CHE stressed for 3000 seconds at V_{GB}=+1.5V, V_{DB}=+5.0V and V_{SB}=0V. No positive oxide charging was observed under this stress condition since the kinetic energy of the hot carriers is only $qV_{GB}-q(V_{DB}-V_{TH})$ = 5.0eV−(1.5eV−0.5eV) = 4.0eV, which is insufficient for the secondary hot holes to surmount the 4.25eV hole barrier at Si/SiO_2 interface [49]. Figure 3.13(a) shows that after the CHE stress, the broad pre-stress structure (V_{GB}<−2.5V) grew almost 200 times in amplitude (to 85pA) and became even broader (V_{GB}<−1V). This indicates a substantial increase of interface trap density in the JSCR. A small post-stress hump can be observed around V_{GB}=−0.5V which was the peak gate voltage due to the pre-stress N_{IT} in the MCR, indicating that some interface traps were also generated in the MCR by the secondary hot holes generated by the hot electrons in the JSCR, but at much smaller magnitude than that in the JSCR due to attenuation of the hot hole density [32].

The pre- and post-stress I_B-V_{GB} data in Fig.3.13(a) can be analyzed using (3.5) to give the lateral distribution of the interface trap density. The results are shown in Fig. 3.13(b). The pre-stress N_{IT} was about $2\times10^{10}cm^{-2}$ from the p/n boundary, y=0, to y=10nm into the JSCR on the basewell or channel side, and it dropped to a nearly constant value of $5\times10^{9}cm^{-2}$ beyond y=16nm. The post-stress N_{IT} increased to $8\times10^{12}cm^{-2}$ within 18nm of the p/n boundary y=0 and dropped continuously towards

the MCR, and is below $10^{12}cm^{-2}$ at y=40nm. The sharp rise near y=0 or the D-LDER in 0<y<3nm may come from the error due to the parabolic potential of the depletion approximation and the resolution limitation of steeply varying $N_{IT}(y)$, as indicated by the theoretical models shown in Figs. 3.4 and 3.5. Thus, this example shows a 3nm spatial resolution.

3.5 Interface Trap Profile in the Channel Region

A single DCIV curve would not provide sufficient information to profile the interface traps in the MCR. Because the surface potential is essentially constant along the entire length of the MCR, every interface trap along the MCR gives the same I_B-V_{GB} dependency in the TE-DCIV measurement regardless of the N_{IT} location. This is readily seen by the general proof using (3.1) for the interface traps in the MCR which reduces to the simple formula:

$$I_B = I_{B1} \int_0^{L_{EL}} N_{IT}(y)Wdy \equiv I_{B1} * N_{ITA}(L_{EL}) \qquad (3.13)$$

$I_{B1}=qR_{SS1}$ is the basewell recombination current from a single interface trap and N_{ITA} is the number of interface traps in the MCR with the electrical channel length of L_{EL} between the edge of source and drain JSCRs. Thus, the spatial variation of $N_{IT}(y)$ is given by

$$N_{IT}(L_{EL})*W = (d/dL_{EL})N_{ITA}(L_{EL}) = (d/dL_{EL})(I_B/I_{B1}) \qquad (3.14)$$

Since L_{EL} is modulated by V_{PN}, the forward bias applied to the Source/Basewell and Drain/Basewell junctions, the V_{PN}-dependent I_{Bpk} measured in MOSTs covering a range of drawn channel length, L, can give the $N_{IT}(y)$ profile near the basewell edge of the two

JSCR's. In addition to TE-DCIV measurement, the SE-DCIV and DE-DCIV measurements would provide further information regarding to the symmetry of the N_{IT} distribution by utilizing the additional variable, i.e., the nearly linear variation of the injected minority carrier interfacial concentration from the emitter, such as the DE, to the collector, such as the zero-biased Source Collector or SC in the DE-DCIV bias configuration. In DE-DCIV, interface traps near the drain-emitter (DE) dominate the I_B while in SE-DCIV, those near the SE dominate. Thus, large asymmetry or differences between DE-DCIV and SE-DCIV would be expected in the stressed transistors and in pre-stressed transistors with designed drain-source asymmetries such as the FLASH memory transistor.

The above differential method will be applied to the 0.25μm family of pMOSTs with 6 drawn channel lengths whose typical TE-DCIV data at 15-V_{PN}'s (0mV to 700mV with 50mV steps) were shown in Fig. 3.10(a) for L=10μm and Fig.10(b) for L=0.25μm. The I_{Bpk}'s were obtained from the LSF to the Gaussian as described by eq. (3.7) with uncertainties of 4-8%. The ratio $I_{Bpk}(L)/I_{Bpk}(L=10\mu m)$ versus V_{PN} for the five drawn L's are plotted in Fig. 3.14(a) and the 4%-8% uncertainties of the datapoints are not perceptible. The nearly V_{PN}-independent value of this ratio is plotted as a function of L in Fig. 3.14(b) for three V_{PN}'s (300, 400 and 500mV) showing two features: **(i)** a rise proportional to L, suggesting an increasing dominance of the N_{IT} in MCR due to increasing length or recombination area, and **(ii)** an asymptotic constant value of 0.25 with decreasing L, suggesting a large contribution to I_B from N_{IT} inside or near the edge of the drain and source JSCRs due to its L-independence. The data points can be fitted

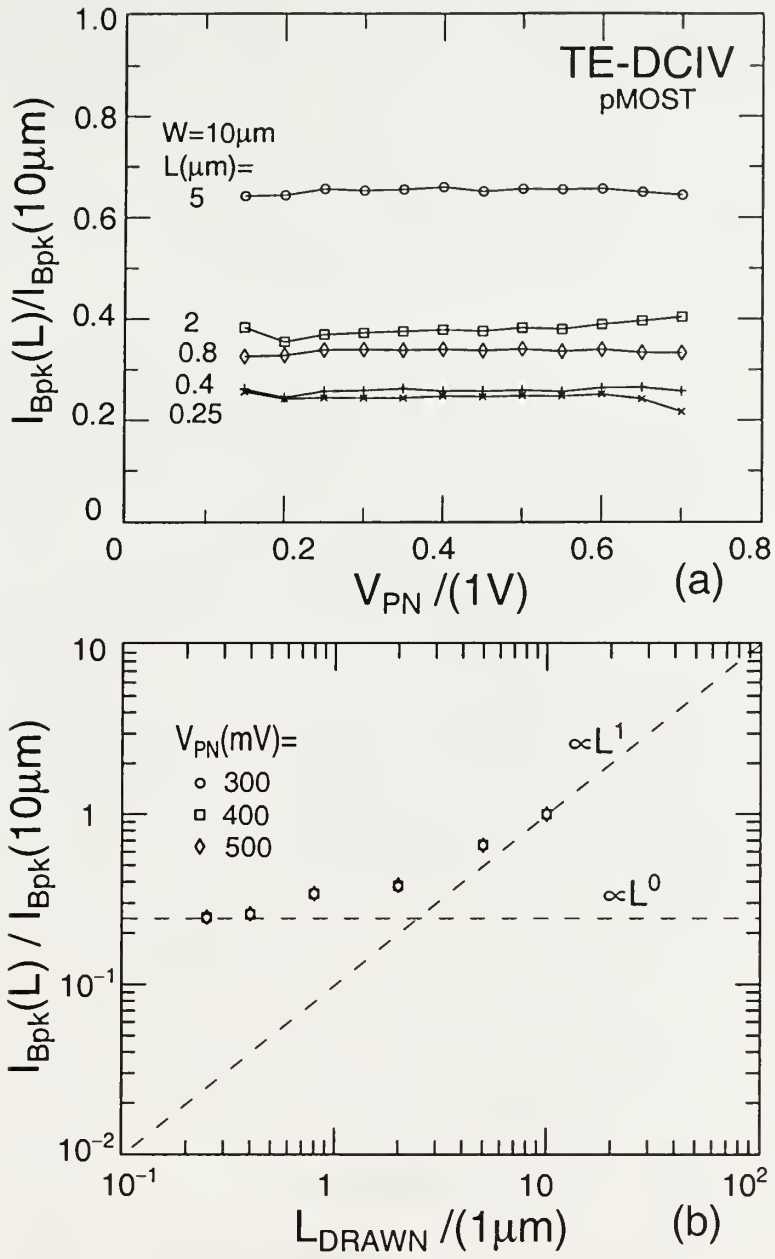

Fig. 3.14 N_{IT} profile extrapolation from channel length dependence of DCIV peak current. **(a)** Normalized I_{Bpk} vs V_{PN} for L(μm)=0.25 to 10. **(b)** Normalized I_{Bpk} vs L at V_{PN}=300, 400, and 500mV. Two components are evident, one each from the MCR and JSCR.

to the following geometric equation where L is Lithographic or drawn channel length which will be taken as the distance between the metallurgical drain and source p/n junction, L_{MCR} is the length of the mid-channel region, $Y_D(\equiv Y_{DJSCR})$ and $Y_S(\equiv Y_{SJSCR})$ are respectively the length of the drain and source junction-space-charge region at the SiO_2/Si interface which are equal, $Y_D+Y_S=2Y_D=Y_J$ for conventional symmetrical cMOS transistor design, and $Y_{MCR} = L - Y_J$, and where the J's are the areal density of the recombination current in the MCR and the drain and source JSCR as indicated by the subscripts which are to be shortened to two characters, MC, DJ, SJ.

$$\frac{I_{Bpk}(L)}{I_{Bpk}(10)} = \frac{J_{MCR}(L - Y_J) + J_{DJSCR}*Y_D + J_{SJSCR}*Y_S}{J_{MCR}(10- Y_J) + J_{DJSCR}*Y_D + J_{SJSCR}*Y_S}$$

$$\equiv y = (x + A_1)/(10 + A_1) \qquad (3.15)$$

where x=L, $A1 = (J_{DJ}*Y_D + J_{SJ}*Y_S)/J_{MC} - Y_J$ which reduces to $A1 = [(J_{DJ}/J_{MC}) - 1]*Y_J$ for identical drain and source. The data in Fig. 3.14(b) gives $A_1 = 3.25\pm0.13\mu m$. For an abrupt source and drain n+/p-B junction of $\sim 10^{18} cm^{-3}$, $Y_D=Y_S=Y_J/2\simeq 0.04\mu m$ giving $(J_{JSCR}/J_{MCR})=N_{IT}(SCR)/N_{IT}(MCR) \sim 40$. This suggests that the $N_{IT}(y)$ profile rises sharply in a short-distance (<0.04μm or <40nm) in the MCR near or inside the JSCR, which was depicted by the theoretical example of a sharper exponential profile, $N_{IT}(y)/N_{IT}(0) = \exp(-y/3nm)$ given in Fig. 3.4(a).

The geometric modeling from the L-dependence just presented that showed a MCR trap and a high-N_{IT} JSCR trap can be analyzed more thoroughly using the I_B integral given by (3.1) to cover the entire channel. In view of the good accuracy and resolution demonstrated by the analytical model of an exponential N_{IT} profile in the

JSCR shown in Fig.4(a)-(c), we shall assume the exponential N_{IT} profile again to model the L-dependence that showed a high-N_{IT} near the edge or inside of the JSCR over the entire V_{PN} range (50mV to 700mV) which sweeps this edge over a region about 20nm to 40nm on the MCR side of the p+D/n-B boundary. Thus, the N_{IT} profile can be modeled by

$$N_{IT}(y) = A_1 + A_2\{\exp(-y/A_3) + \exp[(y-Y_{MCR})/A_3]\} \qquad (3.16)$$

A_1 is the constant concentration in the middle of the channel in asymptotic long channel, A_2 is the peak concentration near the edges of the two JSCR's and $A_3 = L_{IT}$ is the decay length of the N_{IT} distribution.

With the TE-DCIV (both Drain and Source are forward-biased at the same V_{PN}) the minority carrier interfacial concentration, P(x=0,y), in the n-Basewell of a pMOST is spatially constant in the MCR from y=0 to $y=Y_{MCR}$ (y-axis shifted). In the DE-DCIV or SE-DCIV bias configuration, it is linearly decreasing from the forward-biased drain (or source) emitter to the zero-biased source (or drain) collector. These are given below with the origin of the y-axis shifted to the collector-edge of the MCR.

$$\begin{aligned} P(x=0,y) = P_S(y) &= P_S(Y_{MCR}) \qquad \text{TE-DCIV} \\ &= P_S(Y_{MCR})*(y/Y_{MCR}) \qquad \text{DE-DCIV, SE-DCIV} \end{aligned} \qquad (3.17)$$

These three minority carrier concentration profiles can be used with the symmetrical exponential N_{IT} profile of (3.16) to evaluate the I_B integral such as (3.1) with the integration limit extended over the entire channel. To provide the analytical solutions for I_{Bpk}, we shall use the peak value of R_{SS1} in (3.1), which occurs at $c_{ps}P_S = c_{ns}N_S$. For $E_{TI} \sim 0$ or $c_{ps}P_S = c_{ns}N_S >> e_{ps}$ and e_{ns}, it is given by

$$R_{SS1\text{-}pk} = (1/2)(c_{ps}P_s c_{ns}N_S)^{1/2} \qquad (3.18)$$

$$= (1/2)(c_{ps}c_{ns})^{1/2}[n_i \exp(qV_{PN}/2)] \qquad \text{TE-DCIV}$$

$$= (1/2)(c_{ps}c_{ns})^{1/2}[n_i \exp(qV_{PN}/2)](y/Y_{MCR})^{1/2} \qquad \text{DE/SE-DCIV}$$

Then, from (3.1) using (3.16) and (3.18), and L instead Y_{MCR} for the integration limit, and noting that $Y_{MCR} = L - 2Y_{JSCR}(V_{PN})$, we have

$$I_{Bpk} = \{ A_1 L + 2A_2A_3 \cdot [1-\exp(-L/A_3)]\} I_{B1}W \quad \text{TE-DCIV} \qquad (3.19)$$

$$= \{(2/3)A_1 L + \alpha A_2A_3 \cdot [1-\exp(-L/A_3)]\} I_{B1}W \quad \text{DE/SE-DCIV} \qquad (3.20)$$

Where $\alpha \simeq 1$ if $A_3 << L$, or the exponential $N_{IT}(y)$ is localized near the Source and Drain. In the other limit that $N_{IT}(y)$ varies slowly across the channel, $A_3 > L$, $\alpha \simeq 4/3$. These show that I_{Bpk} of TE-DCIV is 1.5 or 2.0 times larger than that of DE-DCIV and SE-DCIV for very long and short channels respectively.

The TE-DCIV, DE-DCIV and SE-DCIV data at V_{PN}=400mV are fitted to (3.19) giving the following results which are used for the LSF lines in Fig. 3.15.

	$A_1 I_{B1}W$(pA/μm)	$A_2 I_{B1}W$(pA/μm)	A_3(nm)
TE	4.3±0.2	90±40	80±40
DE	4.1±0.2	90±50	80±50
SE	4.8±0.2	-	-

These show consistency among the three bias configurations and a $N_{IT}(y)$ that rises from the MCR by a factor of 20 to the edge MCR or JSCR (90/4.2≃21). The two LSF parameters, A_1 and A_2, can be converted to N_{IT}. Using W=10μm, $n_i=10^{10}cm^{-3}$, and assuming $c_{ns}=c_{ps}=10^{-8}cm^3/sec$, we obtained $A_1=2.2x10^9cm^{-2}$ and $A_2=4.68x10^{10}cm^{-2}$ for the interface trap densities in the middle of the channel and near the drain and source junctions.

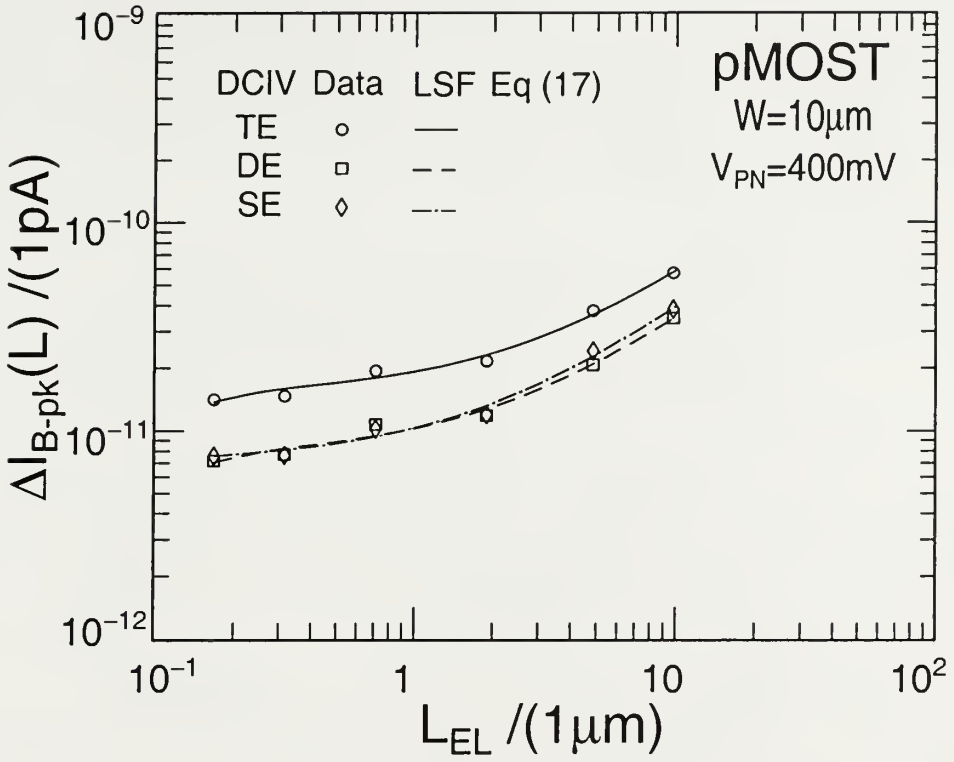

Fig. 3.15 Forward bias configuration dependence of the I_{Bpk}-L data from TE-DCIV, DE-DCIV and SE-DCIV measurements at V_{PN}=400mV, showing the 0.5 ratio at short L from contributions of N_{IT} located inside and near the edge of the ~30nm JSCR and the 0.66 ratio at long L from N_{IT} located in the MCR which are illustrated more explicitly in Fig. 3.16.

The experimental data for the DE/TE and SE/TE ratios of the I_{Bpk} are shown in Fig. 3.16(a) and (b) as a function of forward bias, V_{PN}, showing the predicted variation with channel L by (3.19) from 0.5 to 0.667. Finally, the LSF profile are plotted in Fig. 3.17 at V_{PN}=400mV with an estimated Y_{JSCR}=40nm for the shorter channels, L=0.25μm, 0.4μm, 0.8μm and 2.0μm.

Outside of the MCR or in the JSCRs, N_{IT} seemed to drop as shown by the two broken lines in Fig. 3.17, as suggested by three experimental evidences: (1) The V_{PN}-independence of the I_{Bpk} in Figs. 3.14(a), 3.16(a) and 3.16(b); and channel-length dependence of the (2) lineshape, and (3) $V_{GB\text{-}pk}$ as indicated in Fig. 3.11. The 0.25μm pMOST has a higher shoulder and a larger HWHM at the base-accumulation side of the peak which is consistent with contributions from N_{IT} in the JSCR anticipated by the analytical theory discussed in section 3.3. There is also a 20mV shift of peak gate voltages between the L=10μm and 0.25μm curves which could be from two possibilities: (1) N_{IT} in the JSCR near MCR edge would shift V_{GBpk}(0.25μm) pMOST towards flatband-accumulation (positive V_{GB}) with respect to V_{GBpk}(10μm) and (2) injection-diffusion-limited minority carrier reduction in the 10μm pMOST would shift V_{GBpk}(10μm) towards subthreshold-inversion (negative V_{GB} direction) with respect to V_{GBpk}(0.25μm). The first explanation is favored because N_{IT} inside the JSCR and near the MCR/JSCR boundary can account for the lineshape change as well as peak gate voltage shift as indicated in Fig. 3.11. Furthermore, injection-diffusion-limitation, proportional to L, is expected to dominate at low V_{PN} due to its lower value from the

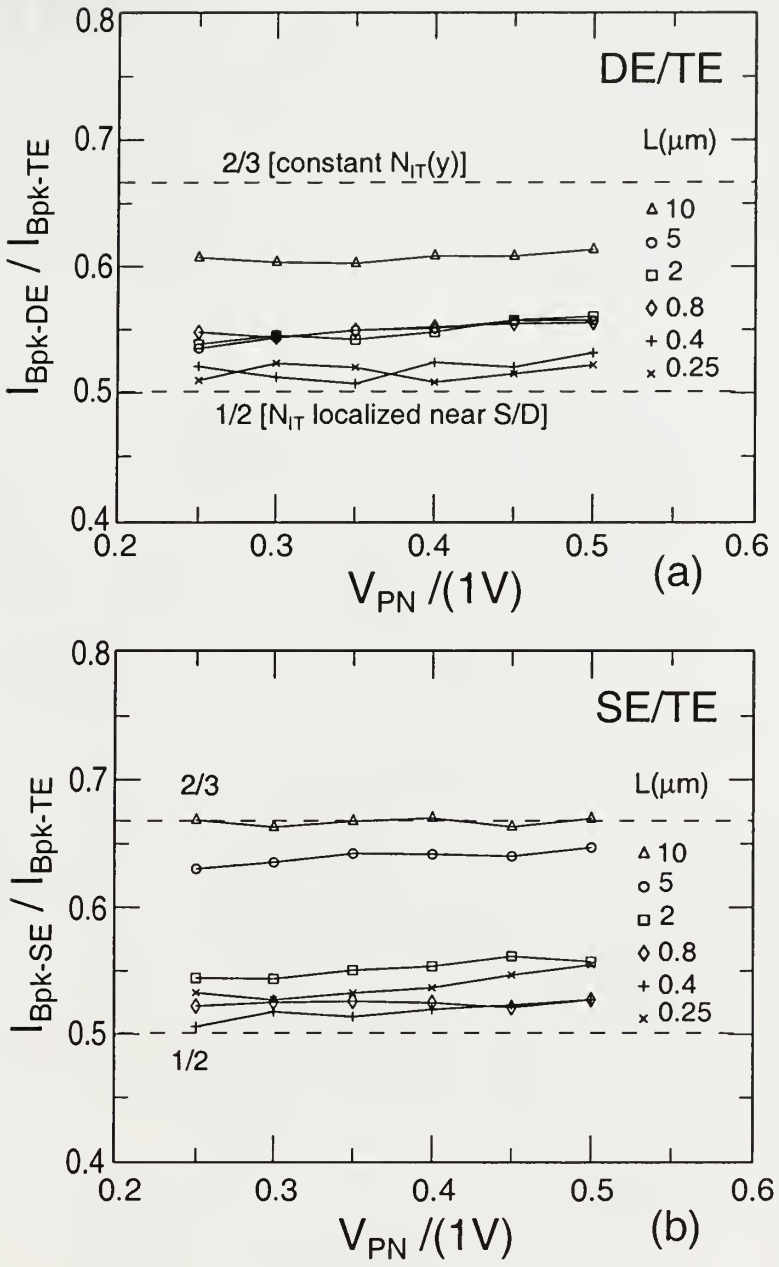

Fig. 3.16 Channel length dependence of the I_{Bpk}-V_{PN} and its theoretical limits for short and long channels. **(a)** DE-I_{Bpk}/TE-I_{Bpk} and **(b)** SE-I_{Bpk}/T_E-I_{Bpk}, showing the theoretically anticipated 0.5 ratio at short L from N_{IT} located inside and near the edge of ~30nm thick JSCR and the 0.66 ratio at long L from N_{IT} located in the MCR.

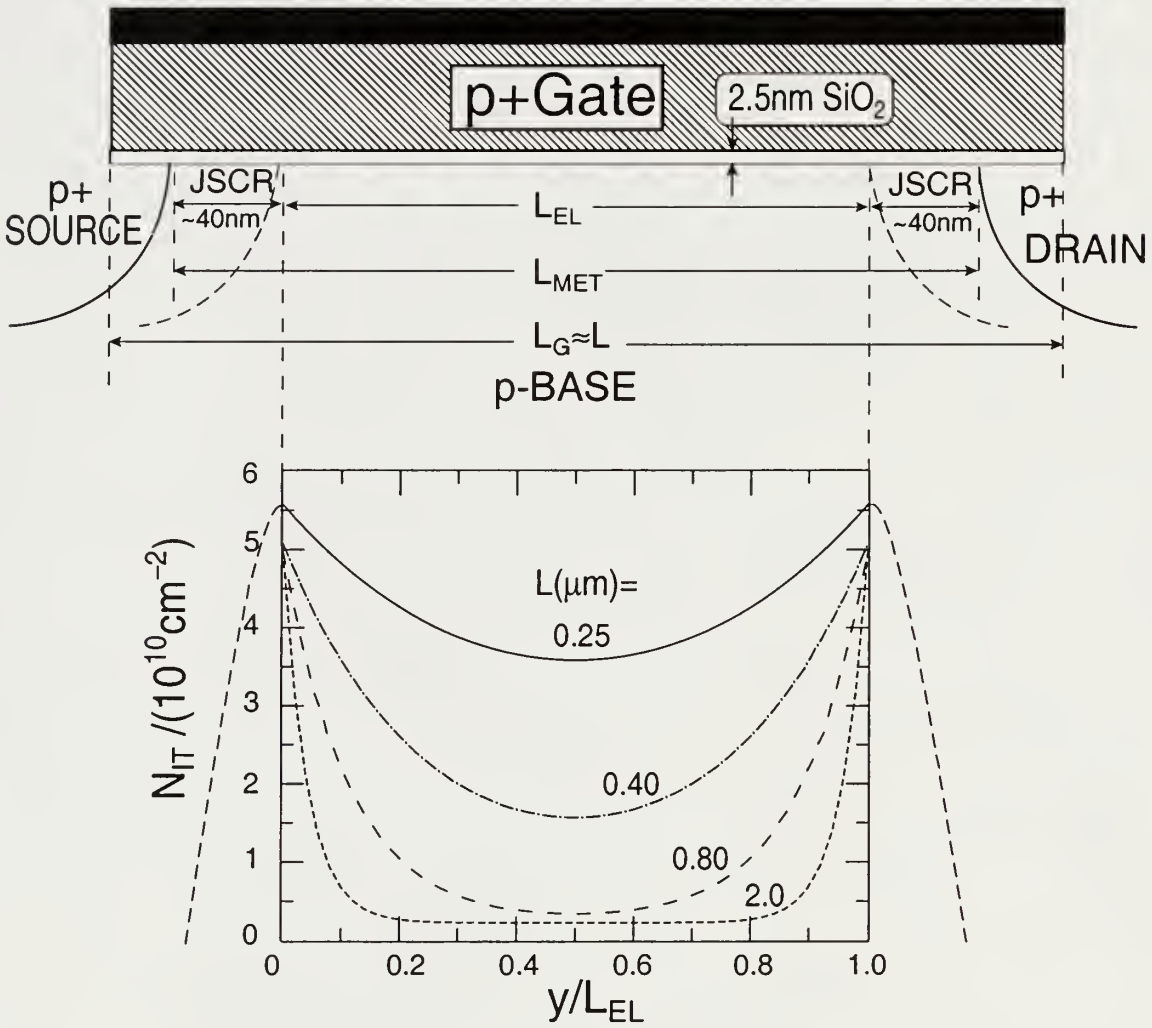

Fig. 3.17 $N_{IT}(y)$ distribution for L(μm)=0.25 to 2.0 of pMOSTs extrapolated from experimental data using the profiling algorithm described in the text.

$\exp(qV_{PN}/kT)$, which is inconsistent with the nearly L-independence of the experimental V_{GBpk} at all V_{PN}'s shown in Fig. 3.12(b).

The presence of the interface traps inside JSCR should not affect the validity of MOST parameter extraction using the V_{PN} dependence of $V_{GB\text{-}pk}$ as described in section 3.4. The parallel shift in Fig. 3.12(b) indicates that the accuracy of oxide thickness and substrate doping concentration extraction would not be affected by the channel length of the MOST. However, there could be a small offset (<20mV) in the extracted flatband voltage.

3.6 Summary

A self-consistent picture on the effects of spatial distributions of interface traps on DCIV characteristics has been obtained and presented. Three applications of the DCIV method have been demonstrated with experimental data: (1) Extraction of transistor design parameters using the weak inversion gate voltage at the peak base terminal current from carrier recombination at interface traps in the channel region; (2) Profiling the concentration of process-residual and channel hot carrier stress generated interface traps in the forward-biased drain junction space charge region using a portion of DCIV curve between the weak inversion gate voltage and strong accumulation gate voltage; and (3) Profiling the concentration of process residual interface traps in the channel region using the channel length dependence of base-current peak. These demonstrations provide further application capabilities for industrial applications of the DCIV method [50,51].

CHAPTER 4
GATE TUNNELING CURRENTS IN ULTRATHIN OXIDE MOS TRANSISTORS

4.1 Historical Survey

Electron tunneling through a classically-forbidden energy barrier was first employed by Fowler and Nordheim [52] (F-N) in 1928 to successfully explain the emission of electrons from the surface of a tungsten electrode into vacuum under the influence of an external electric field of some 10^6 V/cm. Schottky attempted to explain the phenomena by thermionic emission through an image force lowered surface potential barrier [53], but a complete reduction of the 4.5eV barrier in tungsten requires a field of at least 10^8V/cm. In the F-N theory of field emission [54], the electrons inside the tungsten with kinetic energy lower than 4.5eV have an energy dependent probability of tunneling through the field-induced triangular potential barrier as determined by Schrodinger's wave mechanics, while the energy distribution of electrons was obtained following Sommerfeld's electron theory of metal.

With the advancement of quantum theory and materials technology, investigation of tunneling phenomena was extended to (1) extrinsic semiconductors first by Zener [55] which eventually led to the invention of p-n tunnel diode by Esaki [56] in 1957, and (2) metal-oxide-metal (MOM) and metal-oxide-semiconductor (MOS) structures which were made possible by the thermal oxidation technology to grow

surface oxide. In terms of band theories of semiconductors, Kane [57] interpreted the interband tunneling in p-n diode as 'electron penetrating the forbidden gap along the imaginary k axis' and under the constant-field approximation derived an electron tunneling theory that took into account of perpendicular momentum conservation. The p-n tunnel diode showed negative resistance under small forward biases due to the narrowing of energy range for direct interband tunneling with increasing forward-bias but its current never went to zero even when the constant energy tunneling path ceased to exist. This 'excess' current was from many trap-assisted inelastic tunneling pathways, such as those studied by Sah in gold-doped narrow p-n diode [58]. Bardeen considered tunneling in MOM structure as transferring of a quasi-particle across the oxide in a many-particle system [59]. He calculated the transition probability using the time-dependent perturbation approach and showed that the matrix element for the transition is that of the current density operator in the barrier region. This was followed by Harrison [60] to treat tunneling through a forbidden region with spatially varying band structures from an independent-particle point of view. Harrison showed that the traditional Wentzel-Kramers-Brillouin (WKB) approximation can be readily applied to band structures which vary slowly over a wavelength of the particle, but at the sharp boundaries the wavefunctions must be matched to assure the conservation of electron flux. The latter resulted in an energy-dependent pre-exponential factor in the expression of the transition probability which was used by Shewchun and Temple [61] in their computations of tunneling current through semiconductor-insulator-semiconductor strictures. Simmons [62] introduced the average-barrier approximation to evaluate the

WKB tunneling probability for barriers of arbitrary shape and derived an analytical tunneling current formula for MOM structures at zero temperature. Simmon's formula was used by Chang, Stiles and Esaki [63] to correlate their data on electron tunneling through a trapezoidal barrier in a $Al/Al_2O_3/SnTe$ MOS structure while neglecting band bendings in the highly-degenerate p-type SnTe. By making the WKB approximation, detailed tunneling current structures due to the sharp potential boundaries are obscured. Gundlach [64] showed by solving the Schrodinger equation exactly that tunneling current through a triangular barrier can be an oscillatory function of applied voltage due to partial reflections and interferences of electron waves in the conduction band of the insulator at the sharp potential boundary.

Since the advent of silicon transistor technologies in the late 1950's [1], tunneling through metal/SiO_2/Si structures has received much attention. Lenzlinger and Snow [65] performed the original experiment of electron tunneling from silicon into thermally grown SiO_2 and found that the tunneling current density as a function of electric field J(E) followed the classical F-N theory. They derived a constant effective mass of $0.42m_0$ in the forbidden energy gap of SiO_2 and showed that the simple parabolic dispersion relationship is equivalent to that of the empirical Franz-type which reduces to k=0 at both conduction band edge $E=E_C$ and valence band edge $E=E_V$. We note that this equivalence is only valid in the triangular barrier range (or F-N range) while some recent authors applied it to trapezoidal barriers. The band bandings and carrier quantization effects in silicon inversion layer were considered by Weinberg [66,67] and those of silicon accumulation layer considered by Krieger and Swanson [68]

who also included transverse momentum conservation and image force correction in their theory. In view of the accuracy of Weinberg's tunneling J(E) data [66] on thick ~100nm SiO_2 films, we will use them as a reference to calibrate our analytical tunneling theory in the F-N range. On thin ~5.0nm oxide MOS structures, Lewicki and Maserjian [69] observed oscillatory tunneling J(V) characteristics in the F-N range as predicted by the Gundlach theory. From the decay of the oscillation amplitude with the tunneling path length in SiO_2 conduction band, they obtained a value of scattering mean-free-path (MFP) in SiO_2 conduction band, which was later re-interpreted to be 0.65nm by Fischetti et al. [70] who correlated this MFP to the electron-LO phonon scattering. Recently, the oxide fields at which oscillatory extrema occur were used by Zafar et al. [71] to extract oxide thickness in the range of 4.0nm to 6.0nm.

In MOS structures, tunneling carrier species includes electrons and holes, which were first separated by Weinberg et al. who diffused a shallow p-n junction in the silicon substrate under the insulator [72,73]. In a MOS transistor or a MOS capacitor surrounded by a source region (sourced MOS capacitor, or SMOSC), the inversion channel and the well/substrate p-n junction (if available) provide a means for charge-carrier separation. However, this separation is not unambiguous because the carrier type detected are those in the terminals of the p-n junction or the MOST rather than those in the oxide. The following three processes lead to the detection of hole currents as a result of electron tunneling: (1) valence electron tunneling from Si leaves a hole behind (2) electron tunneling from the gate electrode can impact generate a hole or (3) recombine with a hole in the silicon substrate. Careful analysis of charge-separation

data were demonstrated by Yau [74] in his identification of electron as the major conduction carrier in Si_3N_4. Eitan and Kolodny [75] observed a hole current in the p-substrate terminal of a 7.8nm oxide n+sourced MOS capacitor under positive gate bias and attributed it to valence electron tunneling from the p-substrate. This was argued against by Weinberg and Fischetti [76] because their theoretical estimates for valence electron tunneling was about 10^3 times lower than the experimental data. Recently, Rasras et al. [77] used a remote collector to show that this hole current under large positive gate voltage was due to electron-hole pair generation in the p-substrate by photons travelled from the gate electrode. For thin oxide (<3.0nm) under low (<3V) gate voltages, Shanware et al. [78] showed that their valence band tunneling theory with a tunneling effective mass of $0.33m_0$ correlated with the hole substrate currents from which oxide thickness can be extracted.

At the Si/SiO_2 interface, the valence band offset is 4.25eV while the conduction band offset is 3.13eV [79]. The 1.12eV difference in tunneling barrier height makes the hole tunneling much smaller in magnitude and elusive for experimental detection. However, secondary hot hole generation [80] and injection [81] from the anode initiated by F-N tunneling of primary electrons from the cathode was identified to cause positive oxide charge and dielectric breakdown. The threshold SiO_2 voltage to give secondary hole over-the-barrier injection was found by Lu and Sah [81] to be (1) 4.25V by Auger generation and (2) 5V by impact generation of energetic holes. Primary hole tunneling was not reported until very recently in p+gate pMOST [82-84] at the low inversion gate voltage range, but no quantitative correlation with hole tunneling theory was given.

4.2 Recent Technology Motivations

Literatures on fundamental theories and experiments of tunneling are abundant as reviewed in the first section. Most of them focused on tunneling through a triangular barrier. Over the past three decades, semiconductor industry has sustained a growth rate of 15% per year which is fueled by relentless down scaling of silicon transistors. The channel length of a MOST is scaled from 25μm in 1962 [1] to about 0.13μm in the current production technology. Following Dennard's constant-field scaling law [85], gate oxide thickness and power supply voltage need to be scaled accordingly. Oxide thickness is scaled down to about 3.0nm in production technologies while more aggressive scaling to below 1.5nm has been pursued in research and development establishments [86-87] which is limited by oxide tunneling currents. Power supply voltage is limited by the MOST threshold voltage to about 1.0V. At such values of oxide thickness and power supply voltage, the shape of the potential barrier for tunneling is trapezoidal rather than triangle, and gate oxide tunneling could become a limiting factor in MOST off-state power dissipation. Also, tunneling current hinders transistor parameter extraction by traditional measurement techniques such as capacitance-voltage method [10,13,88]. Accurate modeling of trapezoidal tunneling through ultrathin (<3.0nm) oxide is needed in new transistor technology development.

Tunneling through a trapezoidal SiO_2 barrier has been modeled for conduction band electron tunneling from nMOST inversion layers [89-92] and from pMOST accumulation layers [93,94]. When the silicon surface is strongly inverted or strongly

accumulated, electrons at the Si/SiO_2 interface are confined by the surface electric field on the silicon side and the 3.13eV potential barrier on the SiO_2 side. Due to the finite probability of tunneling through the SiO_2 barrier, electrons are in quasi-bound states. Full quantum mechanical models compute the tunneling current by solving the lifetime of the quasi-bound state [89,91,93], which is computationally demanding and not suitable for compact modeling. A comparative study of quantum mechanical and classical modeling, however, reveals similar results in tunneling current-voltage characteristics [93]. This maybe due to the two quantum mechanical effects that act against each other as far as tunneling current is concerned: (1) energy quantization gives a ground state energy above the silicon conduction band minimum which reduces the effective barrier height, and (2) the centroid of electron distribution in the surface layer moves away from the Si/SiO_2 interface which increases 'tunneling distance'. Analytical approximations to evaluate lifetimes [92,94], however, render the tunneling current as a function of oxide electric field and oxide thickness only. Those results give zero tunneling current at zero oxide electric field rather than at the zero gate voltage, which is inaccurate for any MOST structures that has work function difference between gate and substrate materials.

We will develop a compact tunneling current model, which is applicable to the full gate voltage range for both nMOSTs and pMOSTs. The model will treat conduction band electrons, valence band electrons and valence band holes on equal basis. The latter is indispensable to describe the tunneling currents in the inversion range of a p+gate pMOST. Carrier concentrations in silicon layers will be taken into account both

explicitly and implicitly, the latter through the gate-voltage vs. oxide field relationship, and tunneling in both directions will be included. This will allow an accurate comparison of relative importance of various energetic tunneling pathways and geometric tunneling pathways in MOSTs at low gate voltages. We will correlate the model to experimental data and explore its utility for transistor parameter extraction.

4.3 Theory of Tunneling in MOS Structures

In deriving our tunneling current formula, we follow Bardeen's transition probability approach [59] and Harrison's independent-particle tunneling model [60]. The probability per unit time of transition for an electron in state α on one side of the tunneling barrier to state β on the other side is described by Fermi's golden rule:

$$P_{\alpha\beta} = (2\pi/\hbar)|M_{\alpha\beta}|^2 \rho_\beta f_\alpha(1-f_\beta) \qquad (4.1)$$

$M_{\alpha\beta}$ is the matrix element to be evaluated between the two states α and β in the tunneling region where M is proportional to the current density operator as shown by Bardeen [59]. ρ_β is the density of states at β, and f_α and f_β are occupation probabilities of the two states respectively.

The net tunneling current density is derived by summing over all allowed states α and subtracting the reverse tunneling current under the constraint of transverse momentum (k_t) conservation [60]:

$$J = \frac{4\pi q}{\hbar} \sum_{k_t} \int_0^{E_m} |M_{\alpha\beta}|^2 \rho_\alpha \rho_\beta (f_\alpha - f_\beta)\, dE \qquad (4.2)$$

The additional factor 2 accounts for spin degeneracy, q is the charge of an electron and zero to E_m represents allowed energy range for tunneling. The summation over

transverse momentum can be converted to integrals if the electrons are free to move in the transverse directions: $\Sigma \rightarrow 1/(2\pi)^2 \int dk_y \int dk_z$. The double integration over transverse momentum can be carried out explicitly for the Fermi functions (f_α-f_β). The matrix element can be evaluated if the wavefunctions inside the tunneling barrier for state α and β are written in terms of WKB approximations. The result is the familiar exponential WKB tunneling probability. Figure 4.1 shows an energy band diagram of a $Si/SiO_2/Si$ trapezoidal barrier. The Si/SiO_2 conduction band and valence band offsets are denoted by ϕ_{Bn} and ϕ_{Bp} respectively. The initial state on the right side of the barrier with kinetic energy E_1 and wavefunction Ψ_{S1} tunnels to the final state on the left side with kinetic energy E_2 and wavefunction Ψ_{S2}. At the two abrupt boundaries, x=0 and $x=X_{OX}$, both Ψ and $(1/m_s)\partial\Psi/\partial X$ must be continuous to conserve electron flux. Wavefunction matching of Ψ_{S1} at x=0, and Ψ_{S2} at $x=X_{OX}$ results in an electron kinetic energy dependent pre-exponential factor.

Applying the above for electron tunneling through $Si/SiO_2/Si$ structure (see Appendix for detailed analysis), we derive the following tunneling current formula for a single ellipsoid conduction band sub-valley in the silicon:

$$J_N = \frac{4\pi q m_d kT}{h^3} \int_0^{\phi_{Bn}} T_0(E_x) \exp[-2\upsilon(E_x)] \times \log_e \left\{ \frac{1+\exp[(E_F-E_C-E_X)/kT]}{1+\exp[(E_F-E_C-E_X-qV_{GB})/kT]} \right\} dE_X \qquad (4.3)$$

m_d is the density-of-states effective mass of the sub-valley, $E_X=(E-E_C)-E_t$ is the electron kinetic energy in the tunneling direction x, E_F is the Fermi energy, E_C is the

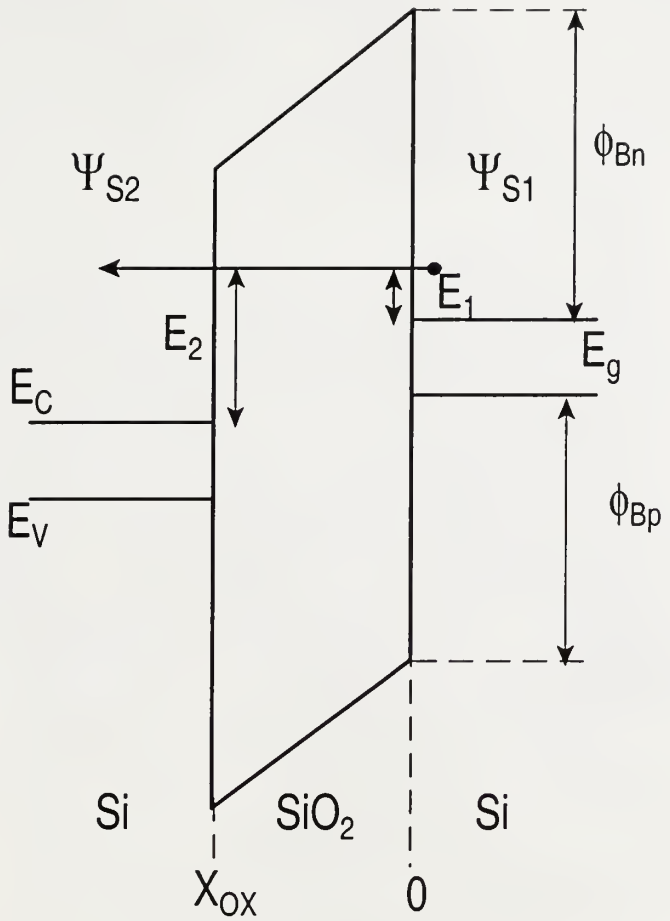

Fig. 4.1 Trapezoidal barrier for electron tunneling in a $Si/SiO_2/Si$ structure. Electron wavefunctions of the initial and final states are denoted by Ψ_{S1} and Ψ_{S2} respectively. Boundary conditions at x=0 and $x=X_{OX}$ must be satisfied to conserve electron flux across the two abrupt boundaries.

conduction band minimum in the semiconductor near the tunneling boundary. $T_0(E_X)$ and $\upsilon(E_X)$ are the pre-exponential factor and the exponent of the WKB tunneling probability respectively to be described as follows in conjunction with figure 4.1.

$$T_0(E_X) = \frac{16}{\left[\frac{m_{ox}k_{1x}(0)}{m_{s1}\kappa_{ox}(0)} + \frac{m_{s1}\kappa_{ox}(0)}{m_{ox}k_{1x}(0)}\right]\left[\frac{m_{ox}k_{2x}(X_{OX})}{m_{s2}\kappa_{ox}(X_{OX})} + \frac{m_{s2}\kappa_{ox}(X_{OX})}{m_{ox}k_{2x}(X_{OX})}\right]} \qquad (4.4)$$

and

$$\upsilon(E_X) = \int_0^{X_{OX}} \kappa(x)\,dx \qquad (4.5)$$

m_{si} (i=1, 2) are the normal effective mass in the two silicon regions, and m_{ox} is the tunneling effective mass in the SiO_2 layer with two boundaries at x=0 and x=X_{OX} respectively. $k_{1x}(0)$ and $k_{2x}(X_{OX})$ are real wave numbers at the silicon side of the boundaries, while $\kappa(0)$ and $\kappa(X_{OX})$ are imaginary wave numbers at the SiO_2 side of the boundaries. Assuming parabolic conduction band of SiO_2 and extending the dispersion relationship analytically into the energy gap, we have the traditional expression for κ, which is:

$$\kappa(x) = [2m_{ox}(\phi_{Bn}-qE_{OX}x-E_X)]^{1/2}/\hbar \qquad (4.6)$$

E_{OX} is the electric field inside SiO_2 and is constant for a trapezoidal barrier. We will include some discussions to the above formulas (4.3) to (4.6) in the following paragraphs.

The WKB approximation is valid when the de Broglie wavelength (λ) of the electron (or its imaginary analogue inside the barrier) is much smaller than the

characteristic length over which the potential varies appreciably [95]. For the Si/SiO_2 electron barrier, λ can be estimated as follows:

$$\lambda = \frac{1}{\kappa} = \frac{\hbar}{(2m_{ox}\phi_B)^{1/2}} \approx 1.5\text{Å} \qquad (4.7)$$

This result shows that the WKB approximation should be valid for oxide thickness as thin as 10-15Å which is already 6-10 times larger than the (imaginary) electron wavelength.

The pre-exponential factor $T_0(E_X)$ of (4.4) is valid for trapezoidal barrier only in which both boundaries are abrupt. This kinetic-energy dependent factor plays an important role in explaining the gate voltage dependence of thin oxide tunnel data as will be shown later. It gives the largest correction near the flatband voltage as compared to the traditional unity pre-exponent. Equation (4.4) is symmetric to the two silicon layers, which is consistent with the notion that the tunneling probability of a trapezoidal barrier is independent of the tunneling direction. In the triangular barrier range, only an approximate form of pre-exponential factor can be derived (see Appendix). A similar factor was used by Krieger and Swanson [68]. As it does not affect the gate voltage dependence of tunneling current, we simply use $T_0=1$ in the triangular barrier range.

The band structures of SiO_2 was studied by Li and Ching [96]. However, the exact dispersion relationship in the wide energy gap of SiO_2 is not known. We use the simplest parabolic band approximation in (6). The empirical Franz-type dispersion relationship has been used by several authors [92,94], which treated the conduction band barrier and valence band barrier on the same footing. However, it assumed equal

effective masses near the valence band edge and near the conduction band edge, i.e., $m_{cb}=m_{vb}$. This is not supported by theoretical band structure calculations [96] that showed a much smaller curvature near the valence band edge ($m_{vb}\geq m_{cb}$). Using the form suggested by Freeman and Dahlke [97], a more appropriate empirical E(k) relationship in the energy gap is:

$$\begin{aligned}\kappa^{-2}(E_X) &= \kappa_{cb}^{-2}(E_X)+\kappa_{vb}^{-2}(E_X) \\ &= \hbar^2/[2m_{cb}(\phi_B-qE_{OX}x-E_X)]+\hbar^2/\{2m_{vb}[E_g-(\phi_B-qE_{OX}x-E_X)]\} \qquad (4.8)\end{aligned}$$

where E_g=8.5eV is the energy gap of SiO_2. Equation (4.8), however, does not yield an analytical solution for the tunneling probability. Since m_{vb} is much larger than m_{cb}, it would be a better approximation by simply dropping the second term on the R.H.S. of (4.8) (the parabolic band approximation) than retaining both terms and making $m_{vb}=m_{cb}$ (Frantz-type dispersion relationship).

It would be straightforward to prove that the equation (3.3) reduces to the classical Fowler-Nordheim formula under these conditions: (1) low temperature: $T\rightarrow 0$, (2) highly degenerate n-type surface layer: $E_F-E_C>kT$ (3) triangular barrier and (4) unity pre-exponential factor: $T_0(E_X)=1$. The classical F-N formula predicts that a semilog plot of J_{FN}/E_{OX}^2 vs. $1/E_{OX}$ is a straight line:

$$J_{FN} = \alpha E_{OX}^2 \exp(-\beta/E_{OX}) \qquad (4.9)$$

$$\alpha = q^3 m_d/(16\pi^2\hbar m_{ox}\phi_B)$$

$$\beta = 4(2m_{ox}\phi_B^3)^{1/2}/(3q\hbar)$$

The total conduction band electron tunneling current is obtained by summing over contributions from all six conduction band subvalleys of silicon. The result

depends on silicon orientation. For <100> Si, there are two longitudinal subvalleys with normal mass m_{si}=0.916 and density-of-state mass m_d=0.190, and four transverse subvalleys with m_{si}=0.190 and m_d=0.417. For <111> Si, there are six equivalent subvalleys with m_{si}=0.258 and m_d=0.358 [98].

Valence electron tunneling and hole tunneling formulas are derived in a similar way. Silicon valence band is consisted of light-hole, heavy-hole and split-off bands. For simplicity, we neglect the contribution from the split-off bands. The minimum barrier height for valence electron tunneling is $\phi_{Bn}+E_g(Si)$=3.13eV+1.12eV=4.25eV, while the maximum barrier height for valence hole tunneling equals the valence band offset ϕ_{Bp}=4.25eV. From each of these valence bands, the tunneling current formula for valence electrons (J_{VE}) and holes (J_P) are listed as follows:

$$J_{VE} = \frac{4\pi q m_d}{h^3} \int_0^{qV_{OX}-E_g} T_0(E_x)\exp[-2\upsilon(E_x)](qV_{OX}-E_g-E_X)\,dE_X \qquad (4.10)$$

$$J_P = \frac{4\pi q m_d kT}{h^3} \int_0^{\phi_{Bp}} T_0(E_x)\exp[-2\upsilon(E_x)] \times \log_e\left\{\frac{1+\exp[(E_V-E_F-E_X)/kT]}{1+\exp[(E_V-E_F-E_X-qV_{GB})/kT]}\right\} dE_X \qquad (4.11)$$

In both formulas, E_X is referenced to the valence band maximum E_V. The valence electron formula has the natural logarithmic term replaced by $qV_{OX}-E_g-E_X$, because of the nearly unity occupation factor of the valence electron state. The upper limit of the integration $qV_{OX}-E_g$ is determined by constant energy tunneling pathways. The crossing of the conduction band and valence band opens the constant energy tunneling

pathways, which determines a threshold oxide voltage for the onset of valence electron tunneling: $V_{OX}=E_g/q$. The hole tunneling formula is almost the same as the conduction band electron tunneling formula except for the band edge reference and the barrier height. The values for the normal mass and density-of-state mass are derived from an expansion of the warped valence band E(k) relationship around the k=0 [99]. They are: $m_{si}=0.28$ and $m_d=0.45$ for the heavy-hole band and $m_{si}=0.20$ and $m_d=0.16$ for the light-hole band.

The only remaining parameter in the three tunneling current formulas (4.3), (4.10) and (4.11) is the oxide tunneling effective mass, m_{ox}, which can be determined by least-squares-fit (LSF) of the conduction band electron tunneling formula (4.3) to Weinberg's data of electron tunneling from <100> Si through a triangular barrier of SiO_2 [66] shown as a solid line in Figure 4.2. Weinberg's F-N tunneling current data was obtained on thick (~100nm) SiO_2 samples under high voltage and therefore has less ambiguity in the values of oxide electric field than the more recent thin oxide data. We fit the theoretical tunneling currents through a MOS structure with thin (5nm) oxide computed from (4.3) to the experimental data in the triangular barrier range at E_{OX}>6MV/cm. The three theoretical curves with n-type impurity doping concentrations from $10^{16}cm^{-3}$ to 10^{20} cm^{-3} converge in this high field range. The tunneling effective mass thus determined is $m_{ox} = 0.437\pm0.005m_0$. This value is used to compute both electron and hole tunneling currents through 2.0nm and 5.0nm oxides. Notice that the ratio of hole currents and electron currents (I_N/I_P) increases as the oxide is thinned or the

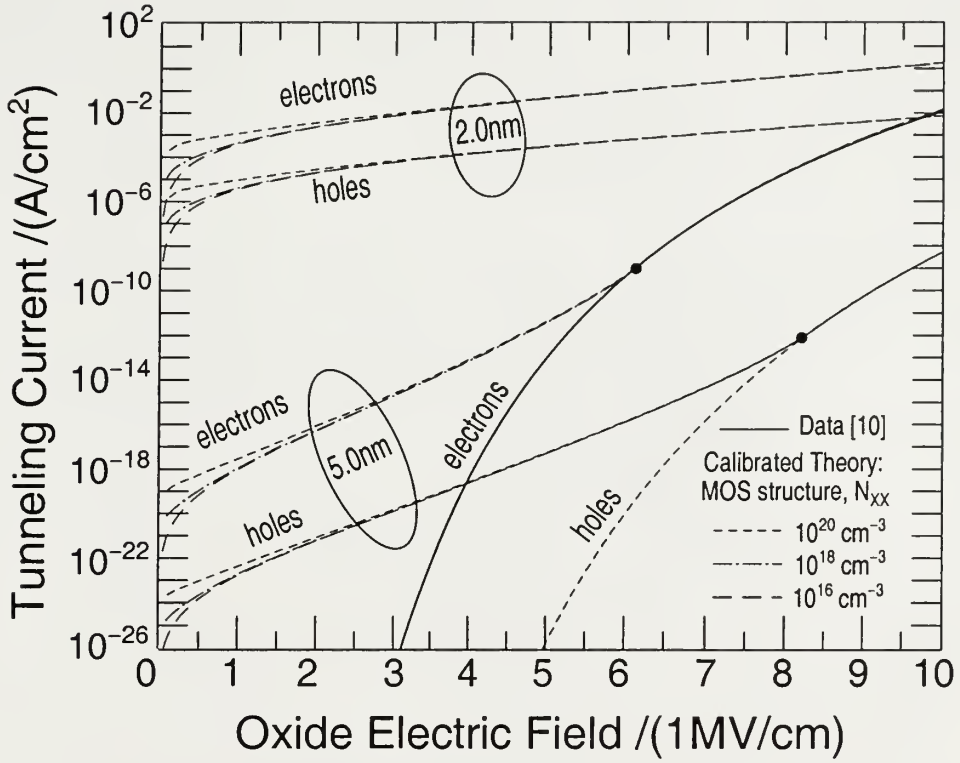

Fig. 4.2 Calibration of tunneling theory to the experimental data (solid line) of electron tunneling from <100> Si through a triangular barrier of SiO_2 in reference [66]. This calibration determines the value of tunneling effective mass which is used to compute the theoretical curves of electron and hole tunneling for two gate oxide thickness (2.0nm and 5.0nm) and three substrate doping concentrations (10^{20} cm^{-3}, 10^{18} cm^{-3} and 10^{16} cm^{-3}).

oxide field is decreased, pointing to the importance of hole currents in thin oxide MOST at low voltages which has not been studied in detail.

4.4 Theoretical Tunneling Current-Voltage Characteristics

We will next compute electron and hole tunneling currents as a function of gate voltage in the trapezoidal barrier range. At each applied terminal voltage, the voltage drops in the gate, oxide and substrate layers are obtained by numerically solving the one-dimensional MOS equation taking into account of Fermi-Dirac distribution and impurity deionization effects. The voltage drops in the gate and substrate layers determine the density-of-states factor at the two interfaces while the oxide voltage together with the oxide thickness determines the barrier transmission probability.

Figure 4.3 shows all three constant energy tunneling current components in two geometrical areas of a p+gate pMOST with 2.0nm gate oxide: over the n-substrate channel area in part (a) and over the gate-overlapped p+source and p+drain areas in part (b). The impurity doping concentrations are those of a typical pMOST: $5\times10^{19}cm^{-3}$ in p+gate, $10^{18}cm^{-3}$ in n-substrate and $10^{19}cm^{-3}$ in p+S and p+D. The three current components in each part are: (1) conduction band electron currents J_N, (2) valence electron currents J_{VE} and (3) hole currents J_P. There are two distinctive features at low voltages. First, hole tunneling current J_P is dominating in both geometric areas which is overtaken by J_{VE} at large negative voltages and by J_N at large positive voltages. Secondly, the overlap area passes much higher hole current density than the channel area. The dominance of hole currents over electron currents is due to the low

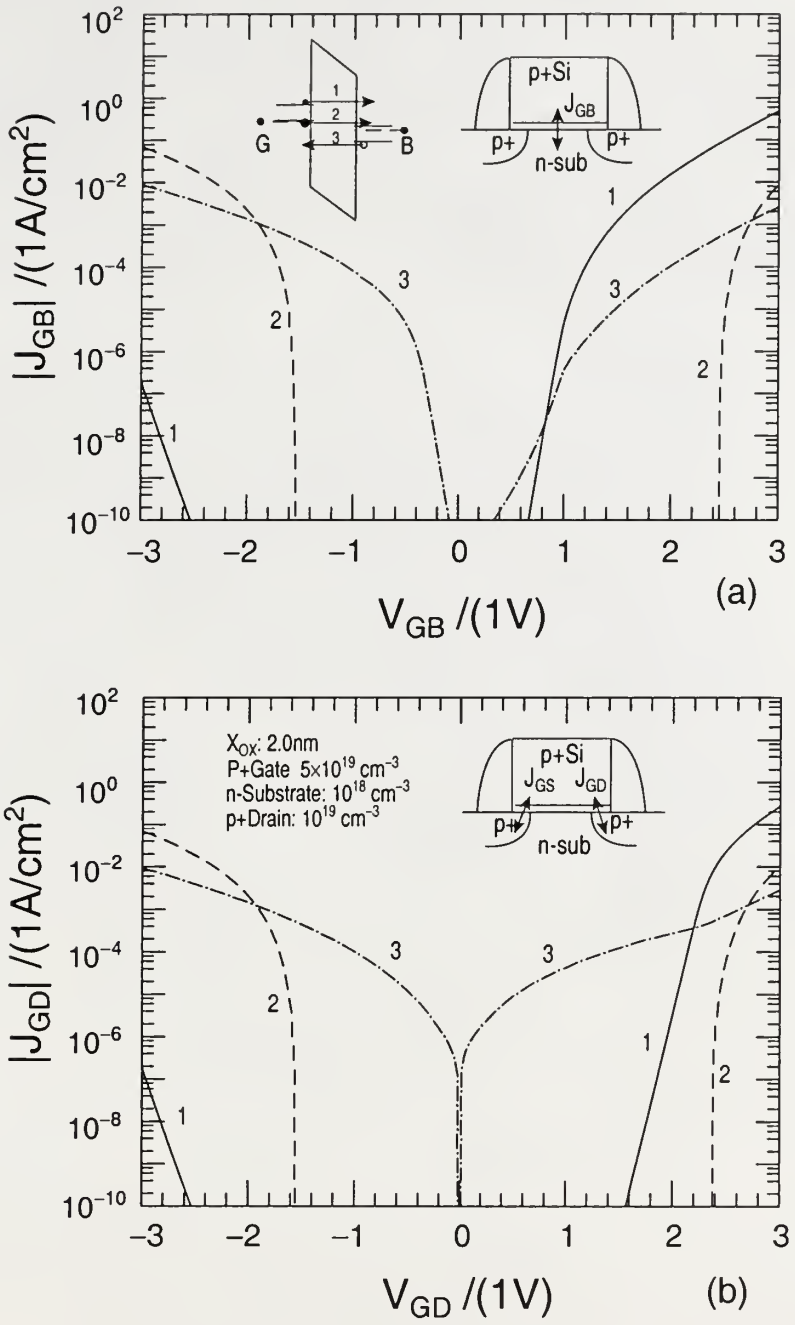

Fig. 4.3 Gate voltage dependence of oxide tunneling current densities for three types of tunneling carriers in pMOST's (**a**) channel area and (**b**) gate-overlapped source and drain areas. The three carrier types are (1) conduction band electrons, (2) holes and (3) valence band electrons. Parameters for the pMOST are: dopant impurity concentrations: $P_{GG}=5\times10^{19}cm^{-3}$, $N_{BB}=10^{18}cm^{-3}$ and $P_{DD}=10^{19}cm^{-3}$, and gate oxide thickness $X_{OX}=2.0nm$.

concentration of conduction band electrons at the interface of p+Si/SiO_2. Tunneling of conduction band electrons from p+Si could become significant only when the surface of the p+Si starts to invert, such as those at $V_{GB}<-2.6V$. The large difference in J_P of the two geometric areas is due to the ~1V difference in gate voltages required for the flatband condition (or $E_{OX}=0$) in the two areas. Flatband occurs at ~1V in the channel area and at ~0V in the overlap area. At $0<V_{GB}<1V$, $V_{OX}<0$ in the channel area which impedes hole tunneling from p+gate, while $V_{OX}>0$ in the overlap area which aids hole tunneling. The negative oxide voltage lowers J_P by raising the integrated tunneling barrier, or υ in (4.5). When the channel is strongly inverted at $V_{GB}<-0.4V$, source and drain areas are electrically connected by the inversion channel and the large difference in J_P of the two geometric areas disappears.

Figure 4.4 compares the three energetic tunneling current components in the two geometric areas in a n+gate nMOST with the complementary impurity type and the same doping concentrations as those of the pMOST in Fig. 4.3. Here conduction band electron current J_N is dominant over the entire voltage range and in both areas as the concentration of conduction band electron is not a limiting factor in n+Gate. The overlap area still passes much higher J_N than the channel area at low voltages for the same reason described in the example of PMOST.

In the next two figures, we illustrate the effects of substrate impurity doping concentrations on the tunneling current-voltage characteristics of a MOS capacitor (MOSC). Figure 4.5 shows J_P in part (a) and J_N in part (b) for p+gate MOSCs, with substrate doping concentration as a parameter which varies from degenerate p-type to

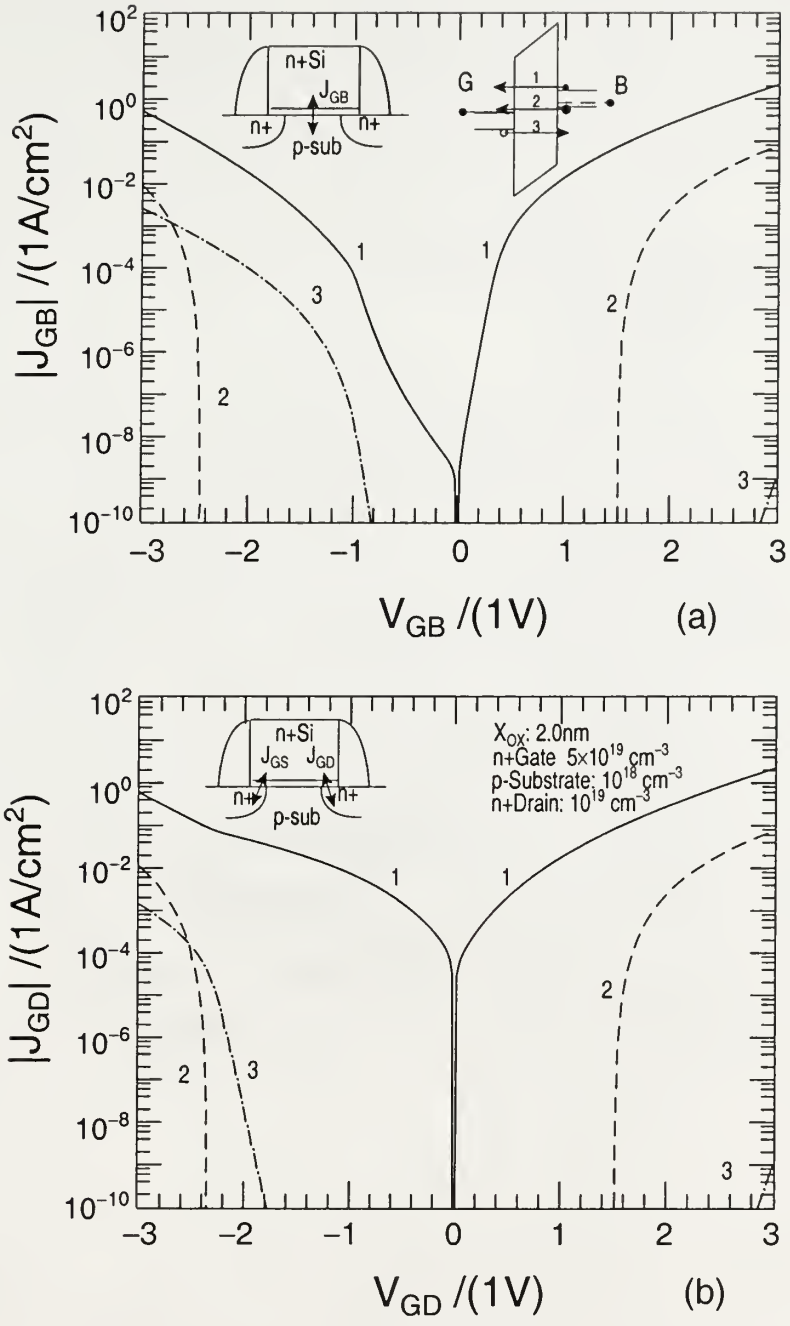

Fig. 4.4 Gate voltage dependence of oxide tunneling current densities for three types of tunneling carriers in nMOST's (**a**) channel area and (**b**) gate-overlapped source and drain areas. The three carrier types are (1) conduction band electrons, (2) holes and (3) valence band electrons. Parameters for the nMOST are: dopant impurity concentrations: $N_{GG}=5\times10^{19}cm^{-3}$, $P_{BB}=10^{18}cm^{-3}$ and $N_{DD}=10^{19}cm^{-3}$, and gate oxide thickness $X_{OX}=2.0nm$.

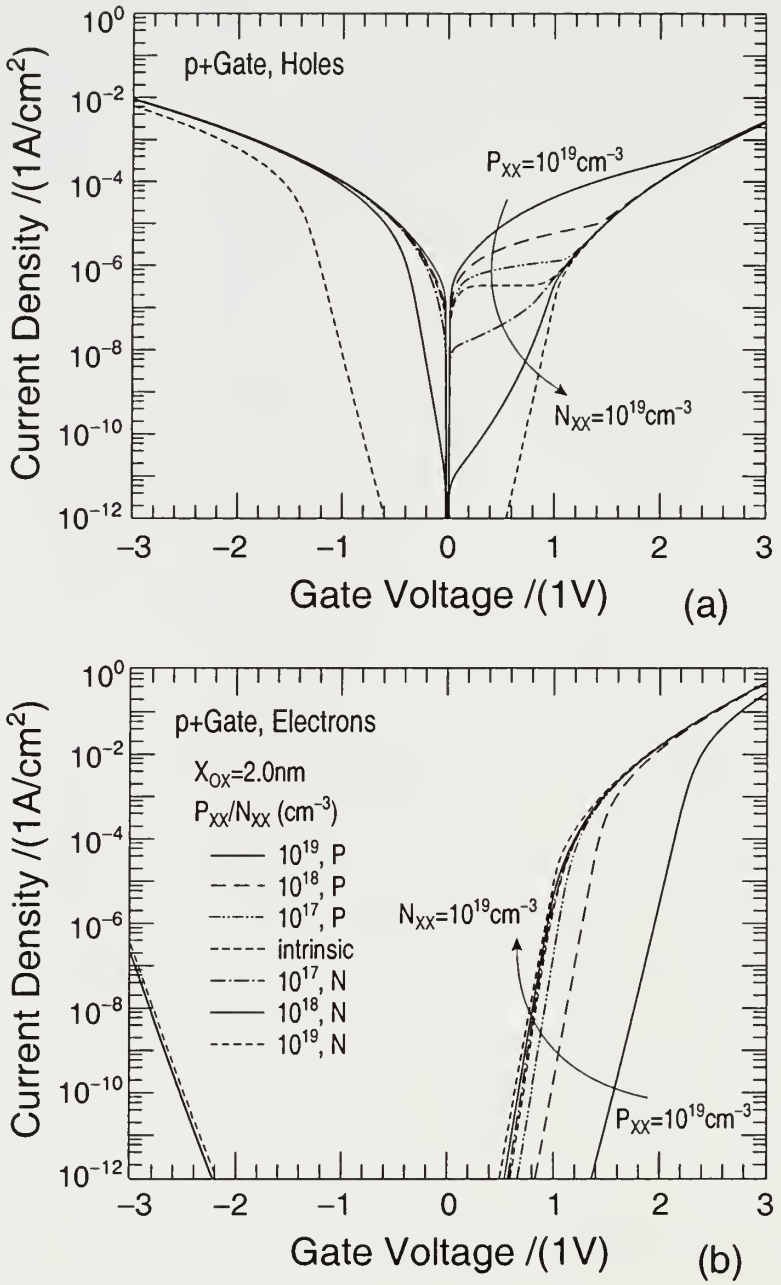

Fig. 4.5 Substrate dopant impurity concentration dependence of gate oxide tunneling currents in p+gate MOS structures. **(a)** Hole tunneling and **(b)** Electron tunneling. $P_{GG}=5\times10^{19}cm^{-3}$, X_{OX}=2.0nm.

intrinsic to degenerate n-type. The general lineshape for J_P-V in (a) and J_N-V in (b) is different, although the oxide voltage is the same for both electron and hole tunneling. This reflects the difference in conduction band electron concentrations at the cathode and valence band hole concentrations at the anode. The most distinctive feature of Fig. 4.5 is the hole current J_P at $0<V_{GB}<1V$ which changes its magnitude and slope with substrate doping concentrations. This feature is attractive for transistor diagnosis as the slope of logJ-V curve is independent of the channel area. At this gate voltage range, holes tunnel from the p+gate to the substrate and J_P is mainly determined by the oxide voltage. For intrinsic substrate, oxide voltage is nearly zero until the Fermi-level in the substrate reaches conduction band edge or valence band edge, this gives a nearly flat J_P. For the three p-type substrates, flatband voltages are close to zero. Higher p-substrate doping concentration results in less voltage drop in the substrate and more voltage drop across the oxide layer which increases the magnitude and slope of the tunneling currents. For the three n-type substrate, however, the flatband voltage V_{FB} is close to 1V and $V_{OX}<0$ at $V_{GB}<V_{FB}$. Therefore larger $|V_{OX}|$ from higher n-substrate doping concentration would reduce the magnitude of tunneling current at $V_{GB}<V_{FB}$. At large positive voltages, when the surface of n-substrate is strongly accumulated or the surface of p-substrate is strongly inverted, the J_P curves would converge as the oxide voltage no longer depends on substrate doping concentrations. This can be understood from the following approximated expressions of oxide voltage drop in the surface inversion range and surface accumulation range of either p+gate/n-substrate MOSC (pMOSC) or n+gate/p-substrate MOSC (nMOSC).

$$V_{OX} = V_{GB}-V_{FB}-V_{SX}-V_{SG}$$

$$\approx V_{GB}-V_{SG} \qquad \text{strong inversion range} \qquad (4.12)$$

$$\approx V_{GB}-E_g/q \qquad \text{strong accumulation range} \qquad (4.13)$$

V_{SX} and V_{SG} are voltage drops in the substrate and gate respectively. The knee points in J_P curves at $V_{GB}<0$ and in J_N curves at $V_{GB}>0$ reflect the condition when the Fermi-level in the substrate reaches the conduction band edge (J_N) or valence band edge (J_P).

Figure 4.6 shows J_N and J_P of n+gate MOSCs for the same variation of substrate dopant type and concentrations. The lineshapes of J_N and J_P are complementary to J_P and J_N of p+gate MOSCs respectively: J_N in Fig. 4.6(a) is nearly a mirror image of J_P in Fig. 4.5(a) while J_P in Fig. 4.6(b) a mirror image of J_N in Fig. 4.5(b). This feature is expected because both the carrier concentration and oxide voltage for J_N (J_P) at V_{GB} in n+gate MOSC are the same as those for J_P (J_N) at $-V_{GB}$ in p+gate MOSC. The magnitude difference is from the ~1eV difference in the barrier height between CB electron tunneling and VB hole tunneling.

Figure 4.7 shows the gate impurity doping concentration dependence of tunneling currents in the two complementary structures: n+gate/p-substrate and p+gate/n-substrate MOSCs. The lineshape of hole currents in part (b) is complementary to that of electron currents in part (a). As the gate doping concentration N_{GG} is increased from $5\times10^{18}cm^{-3}$ to $5\times10^{20}cm^{-3}$ in the n+gate/p-substrate MOSC, electron currents J_N increases but the responsible mechanism varies with the gate voltage range. At $V_{GB}<0$, it is due to increased electron concentration at the n+gate while at $V_{GB}>0$, it

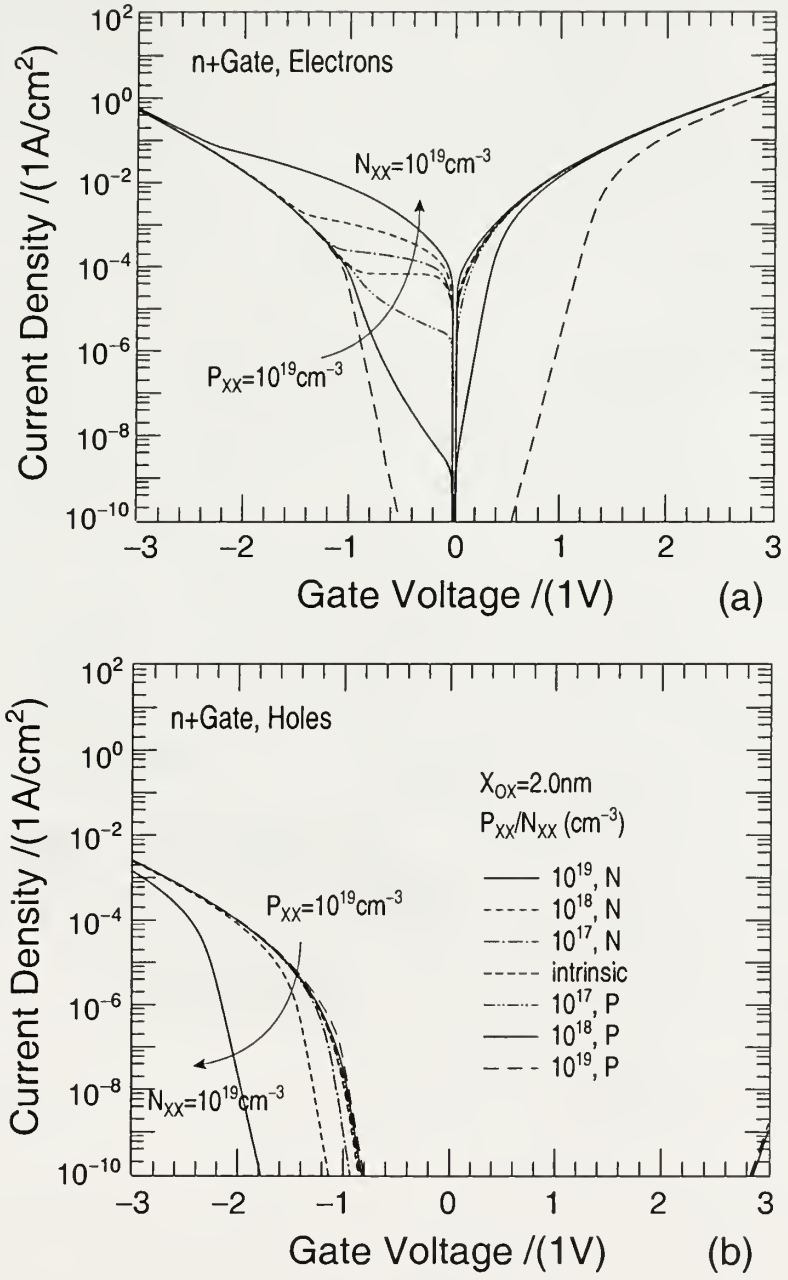

Fig. 4.6 Substrate dopant impurity concentration dependence of gate oxide tunneling currents in n+gate MOS structures. **(a)** Electron tunneling and **(b)** Hole tunneling. $N_{GG}=5\times10^{19}cm^{-3}$, $X_{OX}=2.0nm$.

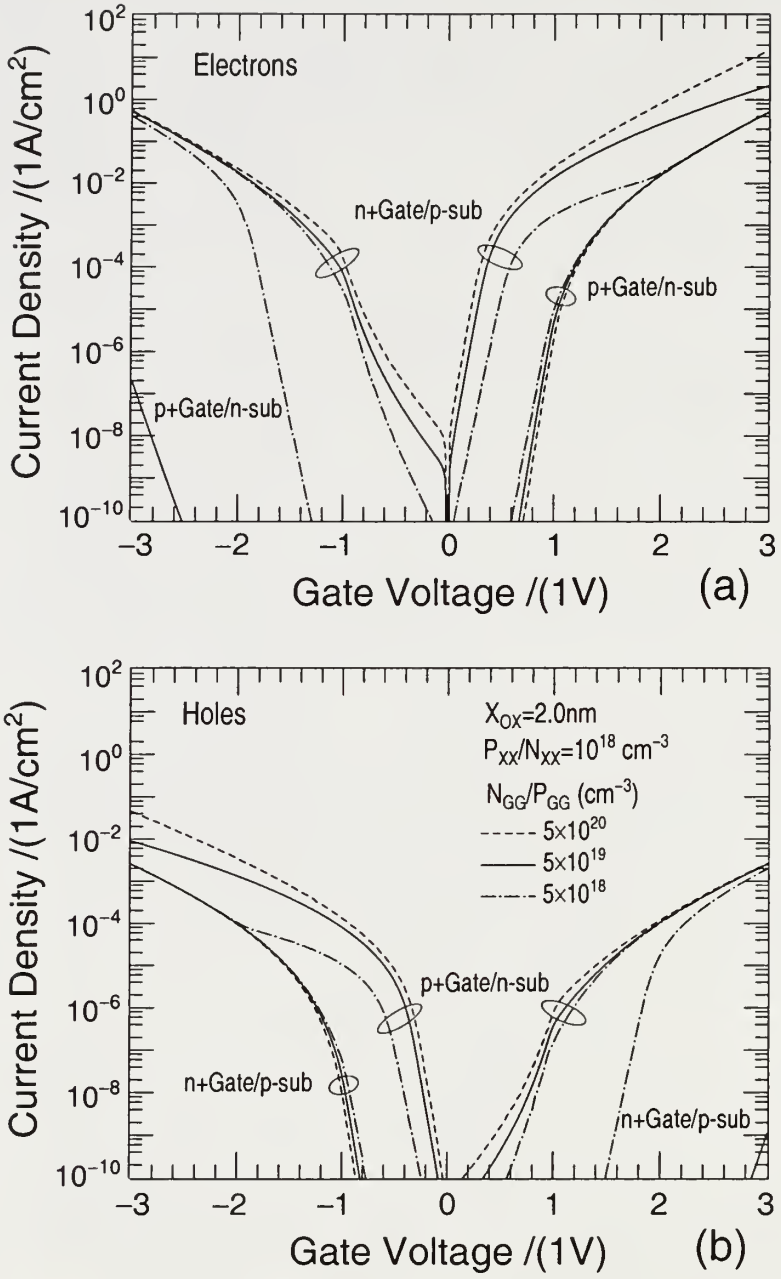

Fig. 4.7 Gate impurity doping concentration dependence of tunneling currents in n+gate/p-sub and p+gate/n-sub MOS structures. **(a)** Electron tunneling currents. **(b)** Hole tunneling currents. Substrate impurity doping concentration is $10^{18}cm^{-3}$ and X_{OX}=2.0nm.

is primarily due to the increase of oxide voltage through the suppression of poly-depletion effect. At large positive voltages, when both n+gate and p-substrate are strongly inverted such as the case of the $N_{GG}=5\times10^{18}cm^{-3}$ curve at $V_{GB}>2.0V$ in part (a), electron currents from the two complementary MOS structures converge. This is because the oxide voltage of the nSMOSC in strong inversion as described by (4.12) approaches that of the pSMOSC in strong accumulation as described by (4.13).

Figure 4.8 shows the oxide thickness dependence of tunneling currents for the two complementary MOSC structures: pMOSC in part (a) and nMOSC in part (b). Hole current is dominant at low voltages in pMOSCs, consistent with the result of Fig. 4.3. In the strong accumulation gate voltage range, the ratio of hole current and electron current increases from 0.001 to 0.1 as the oxide is thinned from 3.0nm to 1.0nm. To our knowledge, hole tunneling theories have not been reported in the literatures. The result of electron tunneling from accumulation layer of a pMOSC (Fig. 14 in [90]) agrees with those of Fig. 4.8(a) at $V_{GB}>1V$, while electron tunneling from inversion layer of a nMOSC (Fig. 8 in [92] and Fig. 15 in [90]) compares well with those of Fig. 4.8(b) at $V_{GB}>0$. For example, the current result gives $J_N=0.23$ A/cm^2 for a nMOS with 2.0nm oxide at $V_{GB}=2.0V$, which is between $J_N=0.15$A/cm^2 in [92] and $J_N=1.0$ A/cm^2 in [90].

Metal gate was used in the early stages of MOS technology development and yet it may be revitalized in the future to replace polysilicon gate in the current MOS technology to eliminate gate depletion effect as well as boron out-diffusion from the p+ polysilicon gate [100]. In figure 4.9, we compare electron tunneling currents through MOSCs with different gate materials including three metals (aluminum, tungsten and

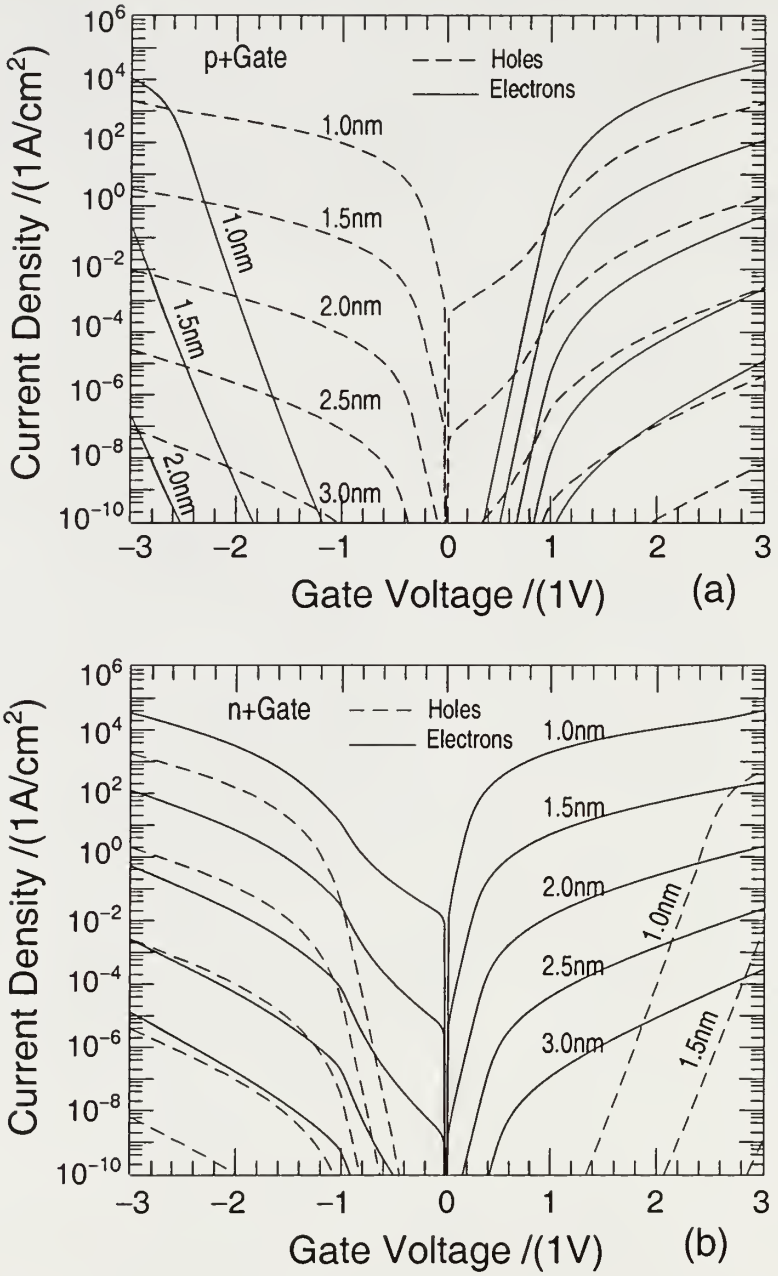

Fig. 4.8 Oxide thickness dependence of electron and hole tunneling currents in MOS structures. **(a)** p+gate/SiO_2/n-substrate. **(b)** n+gate/SiO_2/p-substrate nMOS. Impurity dopant concentrations are $5\times10^{19}cm^{-3}$ in the gate and $10^{18}cm^{-3}$ in the substrate.

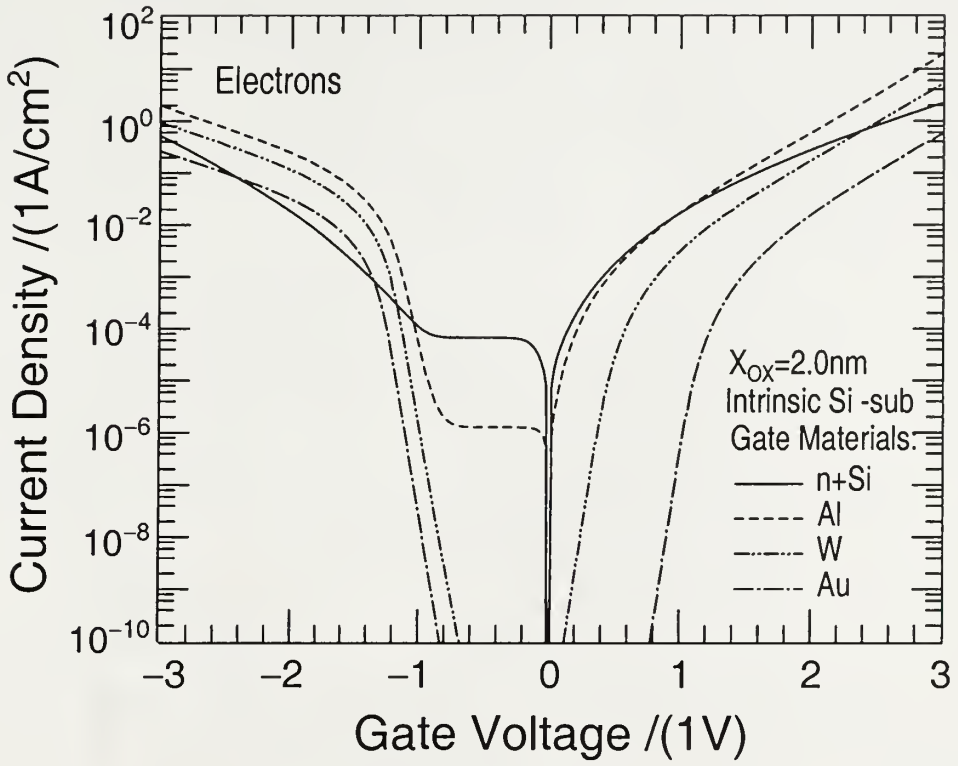

Fig. 4.9 Gate material dependence of electron tunneling currents in MOS structures. Four gate materials include three metals with their respective work functions: aluminum (4.20eV), tungsten (4.57eV) and gold (5.22eV), and silicon with $N_{GG}=5\times10^{19}cm^{-3}$. $X_{OX}=2.0nm$.

gold) in addition to n+ silicon. The work functions of the three metals are: (Al)4.2eV, (W)4.57eV and (Au) 5.22eV [101]. At $V_{GB}>0$, electrons tunnel from the intrinsic silicon substrate. The parallel shift among the three metal-gate curves reflects the work function differences, while the slower rise of the silicon-gate curve is due to silicon gate depletion effect. At $V_{GB}<0$, electrons tunnel from the gate. The work functions determine the tunneling barrier height as well as the oxide voltage and therefore affect tunneling current density in a more complicated manner.

Finally we examine the theoretical temperature dependence of electron tunneling currents in Fig. 4.10. The MOM structure with tungsten gate and tungsten substrate shows very weak temperature dependence, which is expected from our tunneling theory. At $V_{GB}=3V$, J_N only increases about 5% from T=−195°C to T=100°C. Therefore the large temperature dependence at low voltages in MOSC with n+silicon gate and p-substrate must be from semiconductor effects. In particular, the flatband voltage of the MOSC $V_{FB} = -(kT/q)\log_e(N_{GG}P_{XX}/n_i^2)$ shifts positively as the temperature increases due to the $T^{3/2}\exp(-E_g/2kT)$ dependence of intrinsic carrier concentration n_i. The amount of shift is 0.20V from −195°C to 100°C, which changes $V_{OX}(V_{GB})$ in the low voltage range and gives the temperature dependence. At $V_{GB}>0.5V$, tunneling currents at various temperatures converge since the oxide voltage in this strong inversion range as described by (4.12) no longer depends on the flatband voltage.

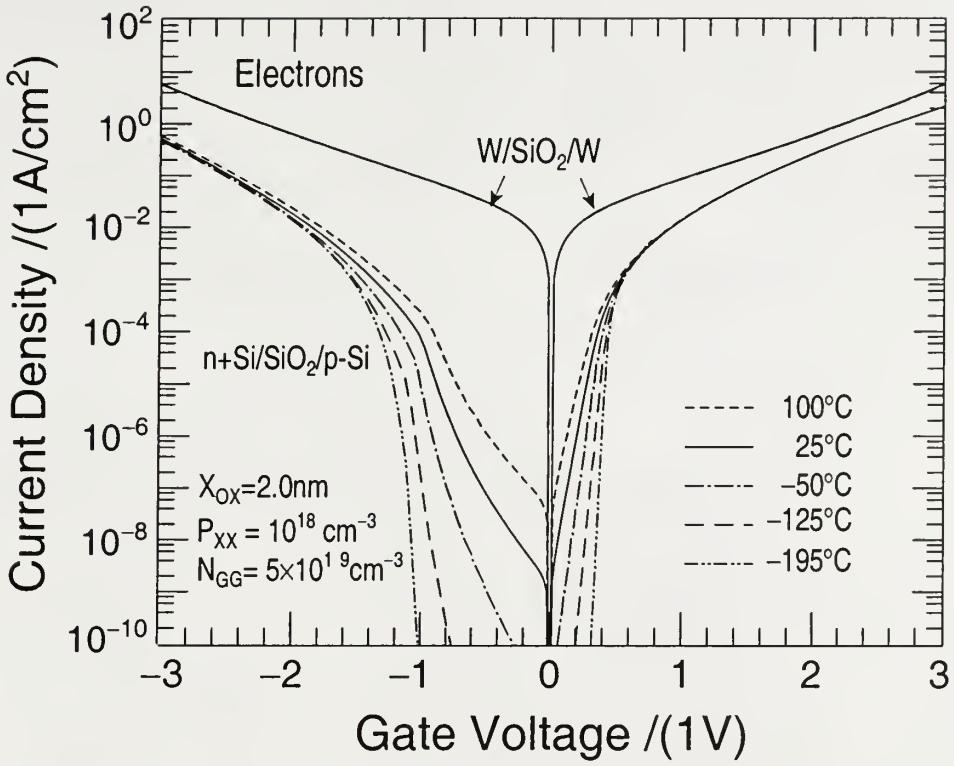

Fig. 4.10 Temperature dependence of electron tunneling currents in MOM ($W/SiO_2/W$) and SOS ($n+Si/SiO_2/p-Si$) structures. Oxide thickness X_{OX}=2.0nm. Temperature varies from 100 degree C to −195 degree C.

4.5 Experimental Correlations

In this section, correlation of the tunneling theory with experimental data on thin oxide sourced MOS capacitors (SMOSC) will be given. These SMOSCs have a diffusion region surrounding the gate area to supply and collect minority carriers when the surface layer of the SMOSC is inverted. The SMOSC are better than a regular MOSC in correlation with the tunneling theory because: (1) The measured terminal currents are not limited by thermal generation of minority carriers in the surface inversion layer of the substrate, (2) Separation of electron and hole currents are made possible by the additional source terminal and (3) It allows the study of the alternative tunneling pathway through the gate-overlapped source region. The same overlap region exists in the drain and source diffusion regions of a MOS transistor so the results on SMOSC can be readily applied to a MOST.

Figure 4.11 shows the typical experimental current-voltage characteristics for two complementary SMOSCs, part (a) for the p+gate pSMOSC and part (b) for the n+gate nSMOSC. The DC bias configurations and the cross-sections for the two SMOSCs are shown on the right side of the experimental data. There are four terminals in a pSMOSC: gate (G), source (S), n-well (B) and p-substrate (C) and only three terminals in a nSMOSC as both the well and substrate are p-type. All terminal currents were measured as a function of the gate voltage between −1V and +1V. The most prominent feature is the complementarity of the gate currents in the two SMOSCs: I_G in the pSMOSC is almost a mirror image of that in the nSMOSC around the axis of

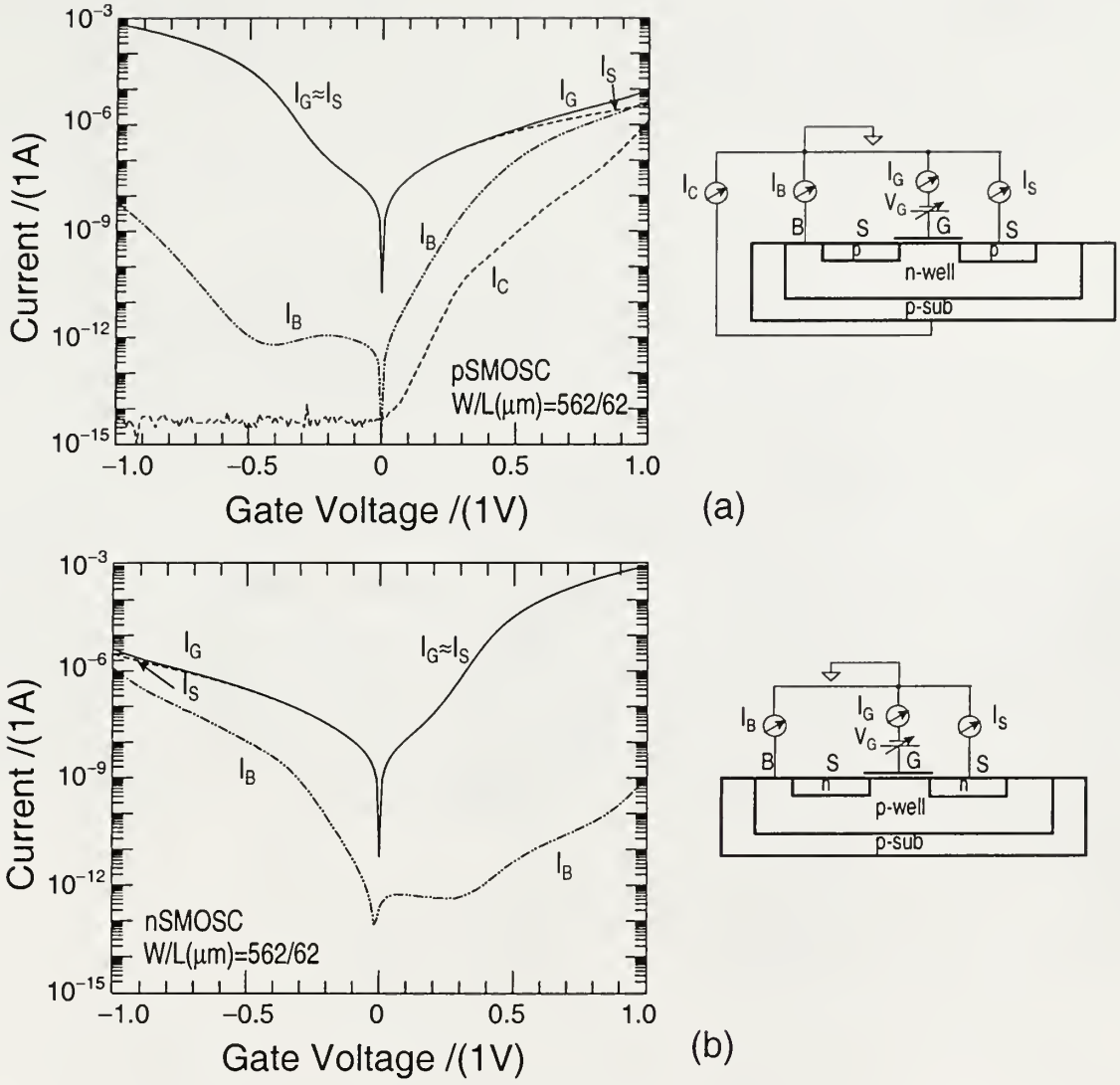

Fig. 4.11 Experimental current-voltage characteristics of sourced MOSCs along with the cross-sections and the DC bias configurations. (a) p+gate SMOSC. (b) n+gate SMOSC.

V_{GB}=0. There is an inflection point in the I_G-V_{GB} curve of the pSMOSC at V_{GB}=−0.3V, which also appears in that of the nSMOSC at V_{GB}=0.3V. The complementarity strongly suggests that if CB electron tunneling is dominant in the nSMOSC then VB hole tunneling must be dominant in the pSMOSC. The oxide voltage is too low in this V_{GB} range for valence electrons to tunnel elastically to the conduction band on the other side of the SiO_2 barrier. However, valence electrons may tunnel to interface traps at the other Si/SiO_2 interface. Next, we will correlate the experimental terminal currents to the three energetic tunneling pathways and the two geometric tunneling pathways, first for the pSMOSCs and then for the nSMOSCs.

In figure 4.12, we breakdown the four-terminal currents in the pSMOSC. Each terminal current is labeled to identify its energetic and geometric tunneling pathways as shown by the cross sections and energy band diagrams below the experimental curves. Pathways 1 (I_G at V_{GB}<−0.25V) and 1' (I_C at V_{GB}>0.3V) are hole tunneling through the channel area (I_{GBp}) from the n-well and from the p+gate respectively, while pathways 2 (I_G=I_S at −0.25V<V_{GB}<0V) and 2' (I_B at V_{GB}>0V) are hole tunneling through the G/S overlap (I_{GSp}) in the two directions respectively. I_B measures electron tunneling current through the channel area but can not be explained by any of the three energetic tunneling pathways as will be shown later. It corresponds to pathways 3 and 4 both of which involve Si/SiO_2 interface traps. A strong evidence for hole tunneling is I_C at V_{GB}>0. The p-substrate collects holes that tunnel from the p+gate and then transverse the n-well without recombination. There is a knee point at 0.3V below which I_C drops faster. This gate voltage corresponds to the onset of weak inversion at the surface of n-well. The

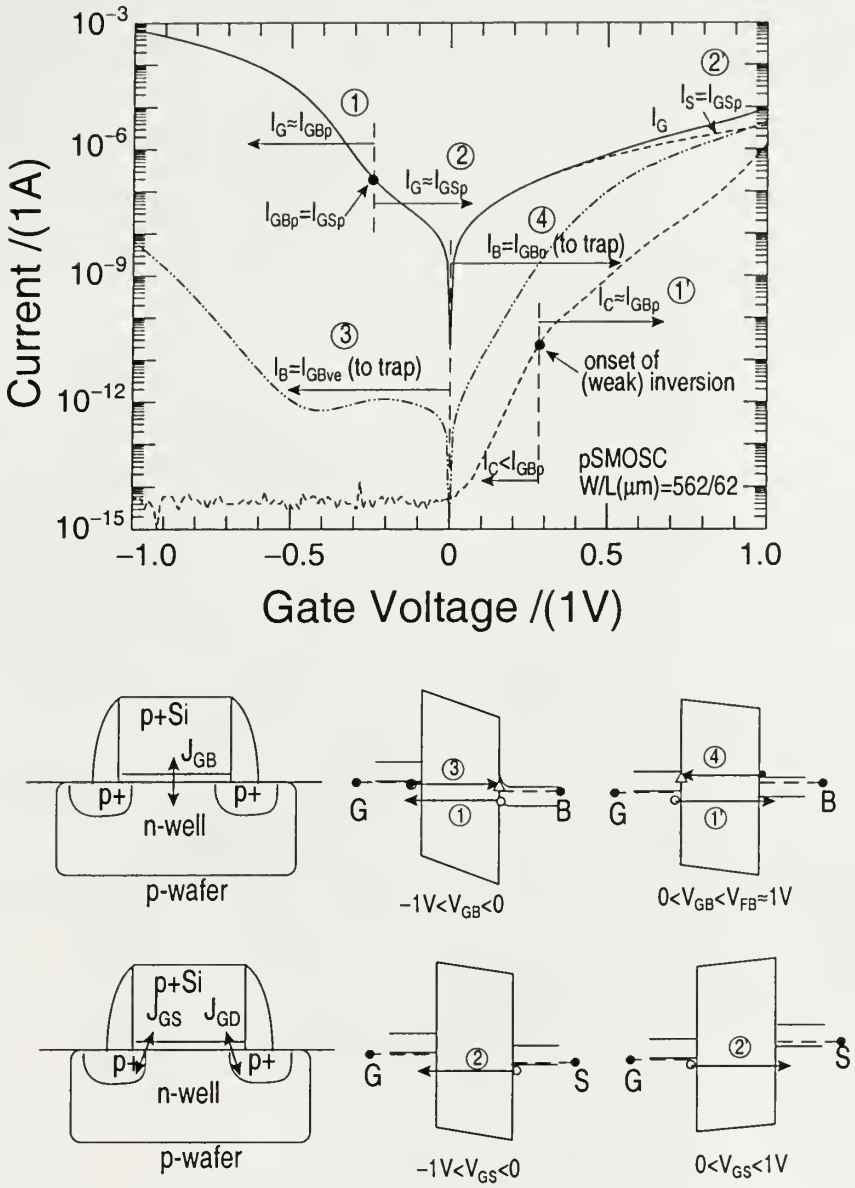

Fig. 4.12 Breakdown of tunneling current pathways in four-terminal (I_G, I_B, I_S, I_C) I-V measurement of a p+sourced p+Si/SiO_2/n-Si MOS capacitor. The labels 1, 1', 2, 2', 3 and 4 indicate the specific energetic and geometric tunneling pathways as shown in the cross-section and energy band diagrams.

inversion channel redirects holes tunneled from p+gate to the source terminal and therefore reducing the hole currents collected by the p-substrate. $I_S \approx I_G$ in the entire V_{GB} range. This is because in the strong inversion gate voltage range, the source terminal supplies the minority carriers (holes) that tunnel through the channel area and in the depletion to flatband gate voltage range, the G/S overlap area is the dominant geometric tunneling pathway. The inflection point in $I_G(I_S)$-V_{GB} curve at −0.25V marks the condition of equal tunneling currents ($I_{GBp}=I_{GSp}$) through the two geometric areas. At $V_{GB}>-0.25V$, I_{GSp} is larger than I_{GBp} despite its much smaller area due to the large difference in the tunneling current densities (J_{GBp} and J_{GSp}) as predicted by the tunneling theory in the previous section, refer to figures 4.3 and 4.5. At $V_{GB}<-0.25V$, I_{GBp} takes over I_{GSp} when the difference between J_{GBp} and J_{GSp} becomes smaller after the surface of n-well is inverted. Next we will quantitatively verify the above analysis using the tunneling theory developed in the previous sections.

Correlation of hole tunneling theory with experimental data on pSMOSC is given in figure 4.13. The three sets of experimental data were measured from three wafers and the variation represents the wafer-to-wafer non-uniformity. In part (a), I_G and I_C are normalized by the channel area $A_{GB}=W\times L=562\mu m\times 62\mu m=3.4844\times 10^{-4}cm^{-2}$. They are matched by the theory of hole tunneling through the channel area (J_{GBp}) computed from (11). The best match was obtained using impurity doping concentrations of $N_{XX}=2\times 10^{18}cm^{-3}$ in the n-well and $P_{GG}=3\text{-}5\times 10^{19}cm^{-3}$ in the p+gate and an oxide thickness of 1.30nm. Each of the transistor parameters affects the tunneling J-V curve in a different way: raising P_{GG} increases the current density in the

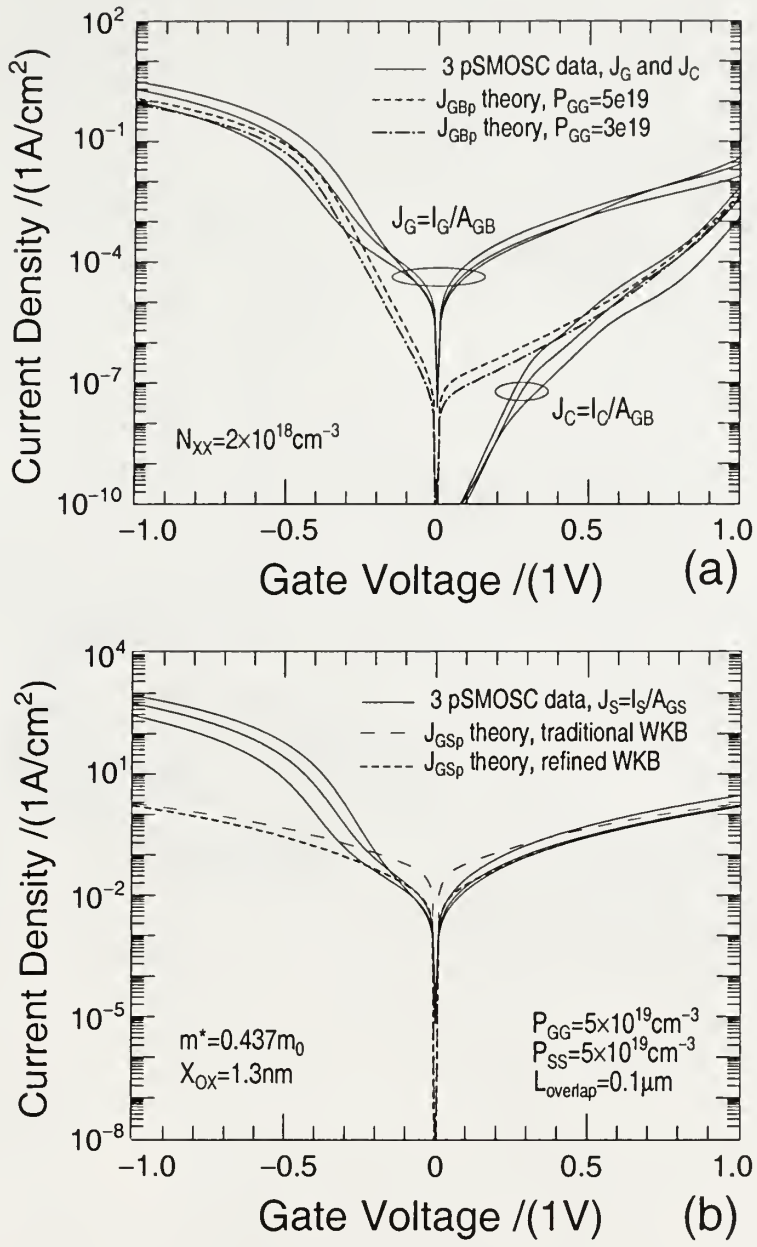

Fig. 4.13 Correlation of hole tunneling theories with experimental data from three p+sourced p+Si/SiO_2/n-Si MOS capacitors. (a) Comparison of J_{GBp} theory with J_G and J_C data. J_G and J_C are the terminal currents I_G and I_C normalized by the channel area respectively. (b) Comparison of two $_{JGSp}$ theories with J_S data. J_S is I_S normalized by the G/S overlap area. The tunneling theories in the two parts use the same transistor parameters including doping concentrations in p+gate ($5\times10^{19}cm^{-3}$), in n-well ($2\times10^{18}cm^{-3}$) and in the p+source ($5\times10^{19}cm^{-3}$) and an oxide thickness of 1.30nm.

strong inversion range near V_{GB}=−1V but does not change J at V_{GB}=1V; raising N_{XX} increases the absolute value of the threshold voltage and shifts the knee point negatively; adjusting oxide thickness gives a vertical shift of the J-V curve. These features can be used judiciously to extract transistor parameters from the tunneling current-voltage characteristics. The agreement between theory and experiment is remarkably well. The theory computed from the same formula matches both J_G data (at V_{GB}<−0.3V) and J_C data (at V_{GB}>0.5V). The deviation of theory and J_C data at V_{GB}<−0.5V was already explained by the surface inversion channel in the n-well that redirects hole currents to the source terminal. The I_G data at V_{GB}>−0.3V is explained in part (b) by the theory of hole tunneling through the G/S overlap area (J_{GSp}). In part (b), I_S (≈I_G) is normalized by the G/S overlap area A_{GS} = 2(W+L)$L_{overlap}$ =2(562+62)μm ×0.1μm = 1.248×$10^{-6}$$cm^{-2}$, assuming 0.1μm for the length of the overlap region. Two theories were computed using the same p+gate doping concentration and oxide thickness as in part (a) and a p+source doping concentration P_{SS}=P_{GG}=5×$10^{19}$$cm^{-3}$. The short-dash line is the result of the current tunneling theory that takes into account of the exact pre-exponential factor of WKB tunneling probability for a trapezoidal barrier, refer to equation (4.4), which is in excellent agreement with the experimental data. The long-dash line is the result of a simplified tunneling theory that uses the traditional WKB tunneling probability with unity pre-exponential factor (T_0=1). The discrepancy between the simplified theory and the experimental data becomes larger at lower |V_{GB}|. This is expected from our previous analysis of the unity pre-exponential factor approximation, the largest error occurs at the flatband condition which is at V_{GB}=0

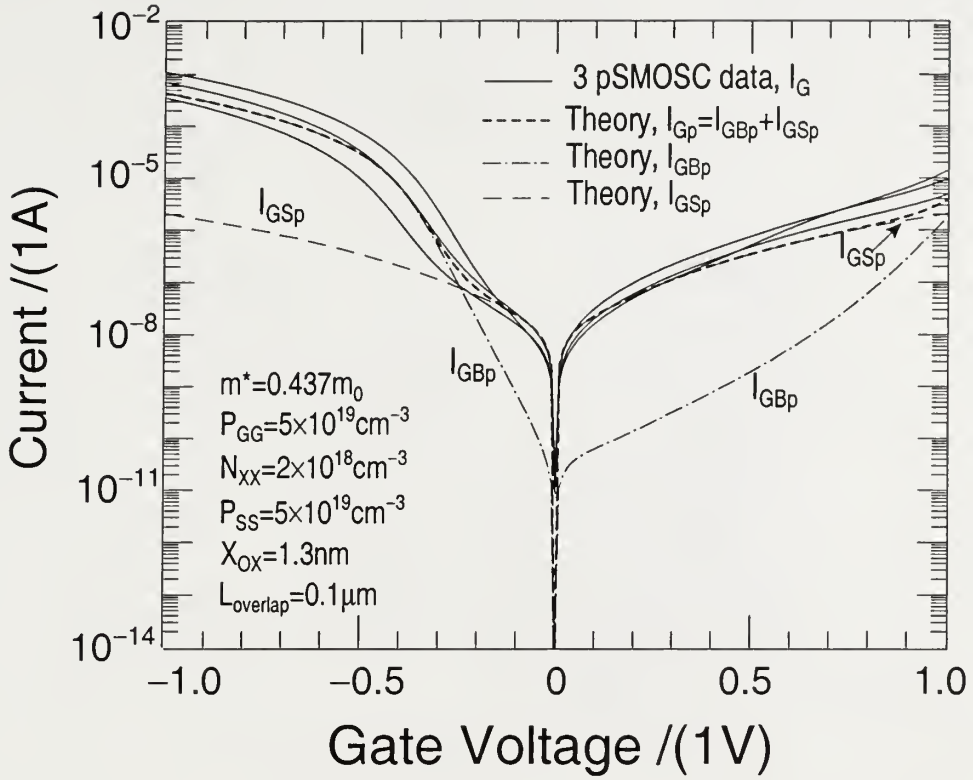

Fig. 4.14 Two tunneling geometrical pathways in p+sourced p+Si/SiO_2/n-Si MOS capacitors. Experimental I_G-V_{GB} curves are matched by adding up two theoretical hole tunneling currents, I_{GBp} for hole tunneling through the channel area, and I_{GSp} for hole tunneling through the G/S overlap area.

for the overlap area. At $V_{GB}<-0.3V$, tunneling current through the channel area starts to dominate and $I_S=I_G$ is already explained in part (a) by the J_{GBp} theory.

In figure 4.14, we add up two theoretical hole tunneling currents, $I_{GBp}=J_{GBp}A_{GB}$ from the channel area and $I_{GSp}=J_{GSp}A_{GS}$ from the G/S overlap area. The result reproduces very closely the experimental I_G-V_{GB} characteristics of the three pSMOSCs. It clearly shows that the inflection point at −0.3V in the experimental data is from the crossing of I_{GBp} and I_{GSp} and that the G/S overlap area is the dominant geometric tunneling pathway in the V_{GB} range from −0.3V to 1.0V despite the area ratio of $A_{GS}/A_{GB}=1/279$! At V_{GB} close to 1.0V, the theoretical hole current can not fully account for the experimental I_G data. This is because the electron tunneling currents as shown by I_B in fig. 4.12 become comparable to the hole tunneling currents. However, the electron tunneling currents in the pSMOSCs can not be explained by any of the trapless electron tunneling pathways, as will be demonstrated in the next figure.

In figure 4.15, we plot the theoretical electron tunneling currents J_{GBn} using equation (4.3) and valence electron tunneling currents J_{GBve} using equation (4.10) for two oxide thickness values of 1.3nm and 1.6nm. Experimental I_B data are much larger than the theoretical predictions for trapless electron tunneling at low voltages. At $V_{GB}>0$, due to thermionic barrier effect, the electron tunneling (J_{GBn}) theory gives a knee point near the flatband voltage ~1V when the conduction band edge E_C on the anode (gate) side lines up with that on the cathode (n-well) side. Each of the experimental I_B curves shows a knee point around 0.4V which indicates electron tunneling from the conduction band of n-well into p+gate/SiO_2 interface traps. This

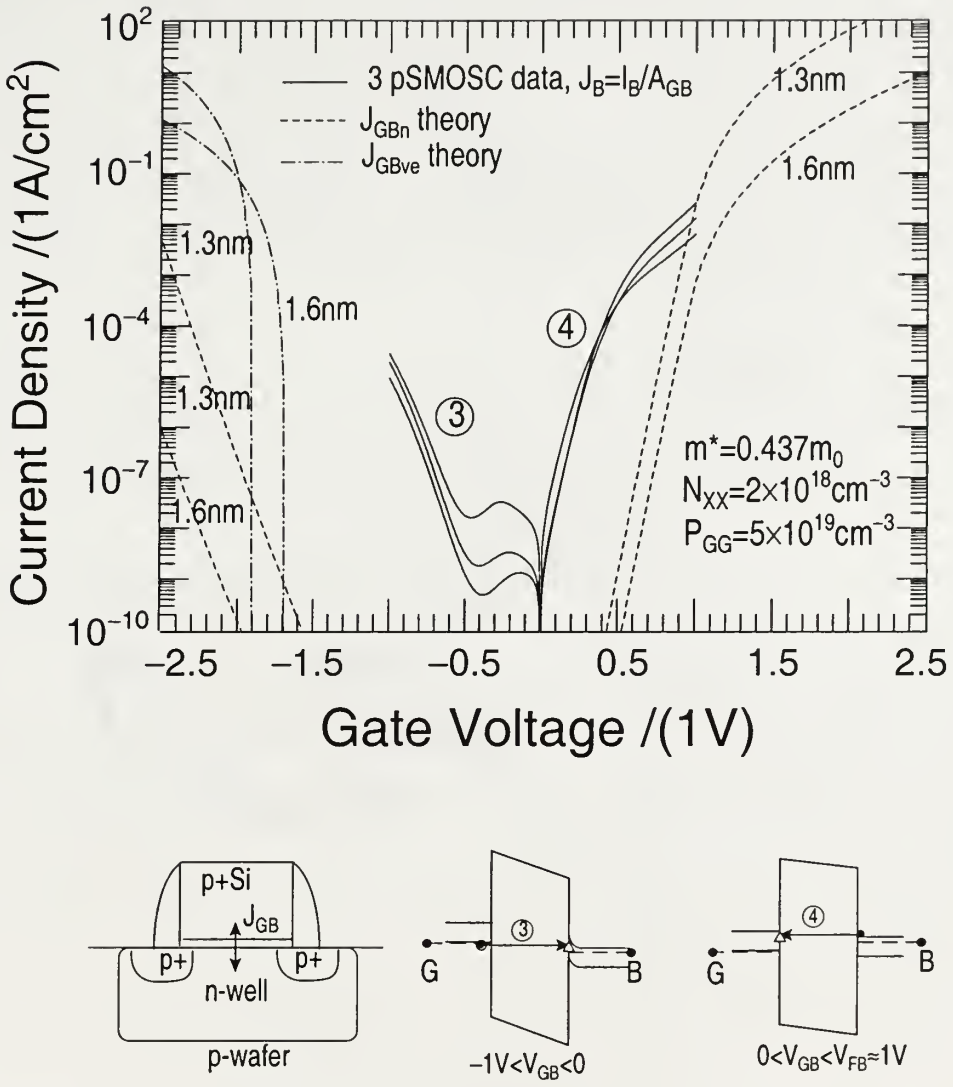

Fig. 4.15 Comparison of experimental data of $J_B=I_B/A_{GB}$ from three pSMOSCs with electron tunneling (J_{GBn}) and valence electron tunneling (J_{GBve}) theories for two oxide thickness values of 1.3nm and 1.6nm. Theory-experimental discrepancy supports the interface-trap assisted tunneling pathway as illustrated by the cross-section and energy band diagrams.

confirms the tunneling pathway number 4 in fig. 4.12, which is also shown in fig. 4.15 below the experimental data. The energy level of the interface trap can be deduced from the oxide voltage drop at the knee point, which is about 0.18eV below E_C. At $V_{GB}<0$, J_{GBn} is much smaller than the experimental I_B data while the trapless valence electron tunneling current J_{GBve} has a threshold $|V_{GB}|$ greater than 1.7V. The theory-experimental discrepancy suggests interface trap assisted valence electron tunneling from the p+gate, which is the tunneling pathway number 3 in fig. 4.12 also shown here in fig. 4.15. I_B is nearly constant at $-0.4V<V_{GB}<0$ and takes off at $V_{GB}<-0.4V$, which corresponds to an oxide voltage of –0.32V in the pSMOSCs. This experimental feature suggests that interface trap energy levels are more than 0.32eV above the valence band edge E_V at the SiO_2/n-well interface.

Following the above approaches for the pSMOSCs, we will next give the theory-experimental correlations for the nSMOSCs. The nSMOSC data in Fig. 4.11(b) is analyzed in detail in Fig. 16 to identify the specific energetic and geometric tunneling pathways in each of the three terminal currents I_G, I_B and I_S. In the nSMOSC, conduction band electron is the dominant tunneling carrier species. Electron tunneling through the channel area I_{GBn} is identified for $I_G=I_S$ at $V_{GB}>0.25V$ (pathway 1) and I_B at $V_{GB}<-0.35V$ (pathway 1'), while electron tunneling through the G/S overlap area I_{GSn} (pathways 2 and 2') is identified for $I_G\approx I_S$ at $V_{GB}<0.25V$. Similar to the pSMOSCs, the inflection point in the I_G curve at $V_{GB}=0.25V$ is due to the crossing of the two geometric electron tunneling components, i.e., $I_{GBn}=I_{GSn}$, and the knee point in the I_B curve is due to the onset of (weak) surface inversion in the p-well. All these will be quantitatively

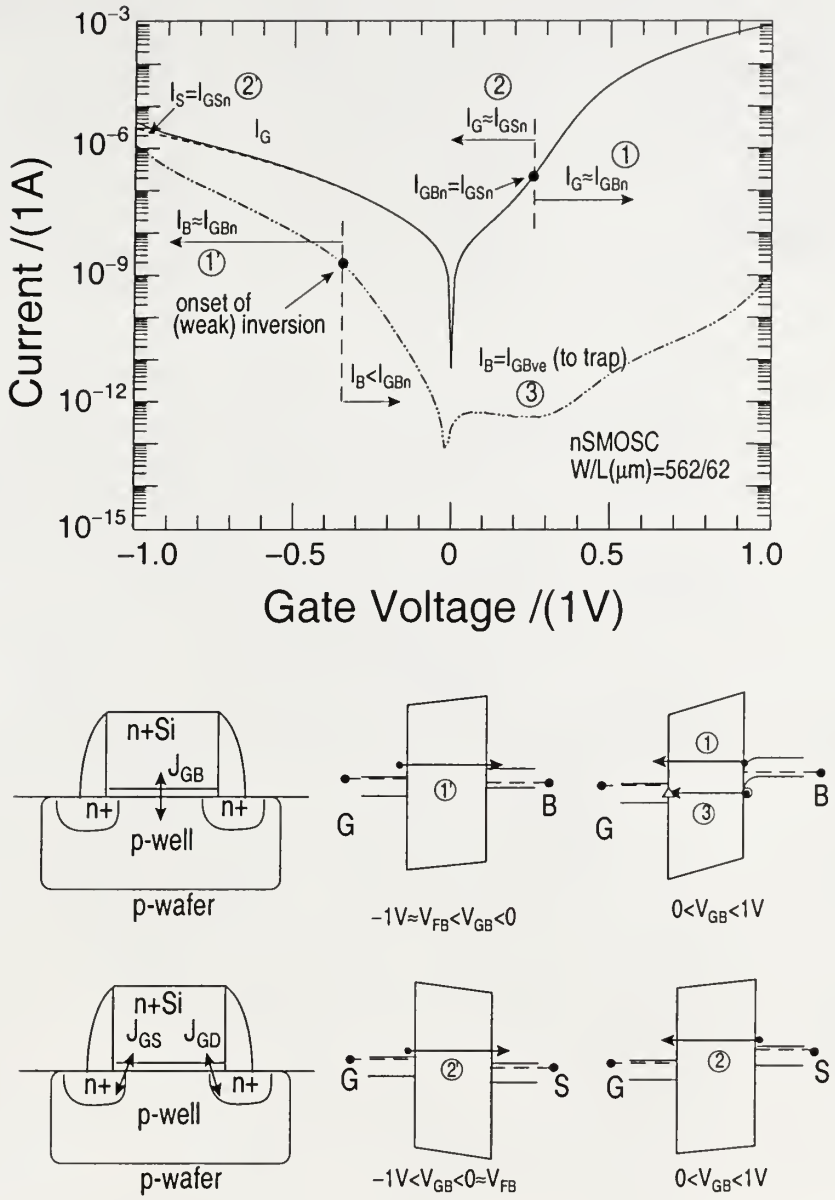

Fig. 4.16 Breakdown of tunneling current pathways in three-terminal (I_G, I_B, I_S) I-V measurement of a n+sourced n+Si/SiO_2/p-Si MOS capacitor. The labels 1, 1', 2, 2' and 3 indicate the specific energetic and geometric tunneling pathways as shown by the cross-section and energy band diagrams.

verified by the tunneling theory. However, the trapless tunneling theory can not explain I_B in the positive gate voltage range, which is likely due to valence electron tunneling from the p-well to interface traps at the n+gate/SiO_2 interface. All the electron tunneling pathways in the nSMOSC are further illustrated by the cross-sections and energy band diagrams in figure 4.16.

Two sets of nSMOSC data that have different I_G-V_{GB} lineshapes will be used to correlate with the tunneling theory. The second set of data are from a previous development technology with lower p-well doping concentrations.

Figure 4.17 gives the theory-experimental correlation for three nSMOSCs from a recent technology which includes the nSMOSC shown in Figs. 4.11 and 4.16. In part (a), theory of electron tunneling through the channel area (J_{GBn}) is compared to the experimental data of $J_G=I_G/A_{GB}$ and $J_B=I_B/A_{GB}$. The best matches shown here are obtained by using impurity doping concentrations of $5\times10^{19}cm^{-3}$ for the n+ gate and $1\text{-}2\times10^{18}cm^{-3}$ for the p-well and an oxide thickness of 1.60nm. The three theoretical curves correspond to p-well impurity doping concentrations ($\times10^{18}cm^{-3}$) of 1.0, 1.4 and 2.0 respectively. The first curve (1.0) fits the depletion-flatband range ($V_{GB}<-0.3V$) well, the third curve (2.0) fits the inversion range ($V_{GB}>0.5V$) well, while the second curve (1.4) gives the best overall fit. The faster drop of experimental J_B data towards zero in the negative V_{GB} range is expected because of the surface inversion channel that redirects electrons tunneled from the n+gate to the source terminal, which reduces the number of electrons being injected into the p-well/p-substrate. I_G at $V_{GB}<0.3V$ is dominated by electron tunneling through the G/S overlap region, as proved by theory-

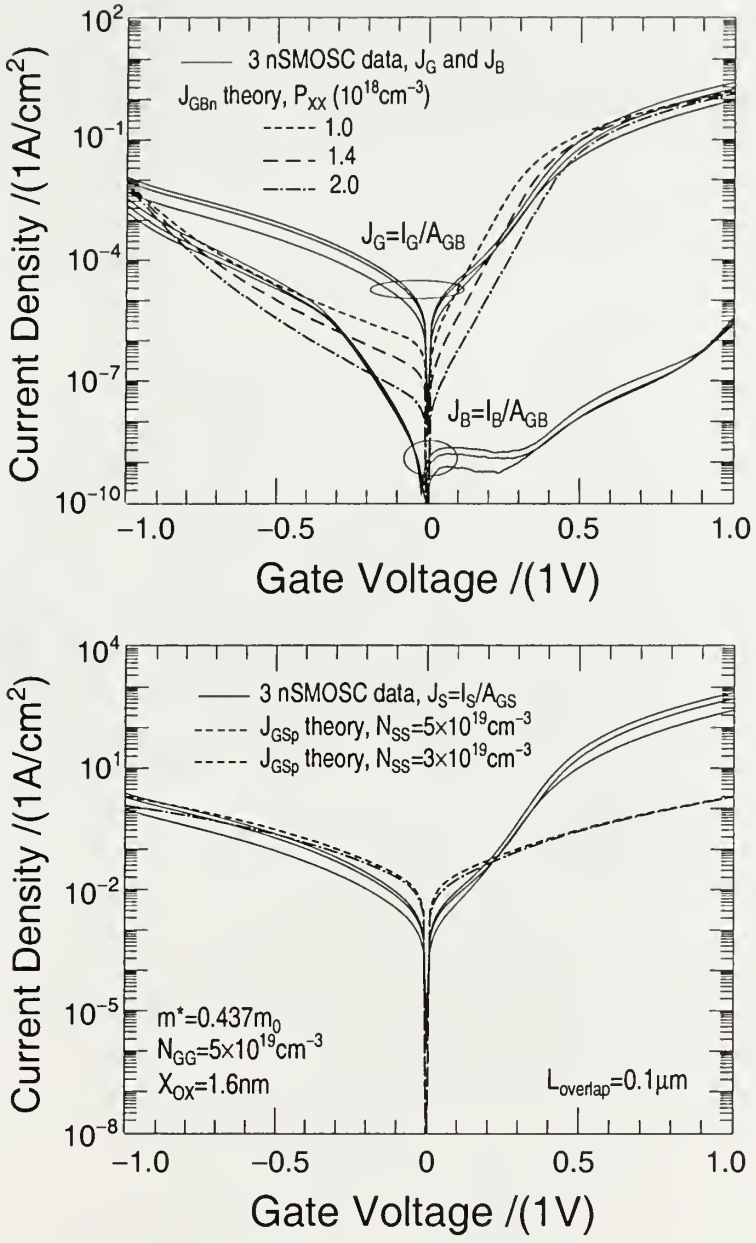

Fig. 4.17 Correlation of electron tunneling theories with experimental data from three n+sourced n+Si/SiO_2/p-Si MOS capacitors. (a) Comparison of the J_{GBn} theory with $J_G=I_G/A_{GB}$ and $J_B=I_B/A_{GB}$. (b) Comparison of the J_{GSn} theory with $J_S=I_S/A_{GS}$. The tunneling theories in the two parts use the same n+gate doping concentration of $5\times10^{19}cm^{-3}$ and oxide thickness of 1.60nm. Three theoretical J_{GBn} curves in part (a) correspond to p-well doping concentrations ($\times10^{18}cm^{-3}$) of 1.0, 1.4 and 2.0 respectively, and the two theoretical J_{GSn} curves in part (b) corresponds to n+gate doping concentrations ($\times10^{19}cm^{-3}$) of 3.0 and 5.0 respectively.

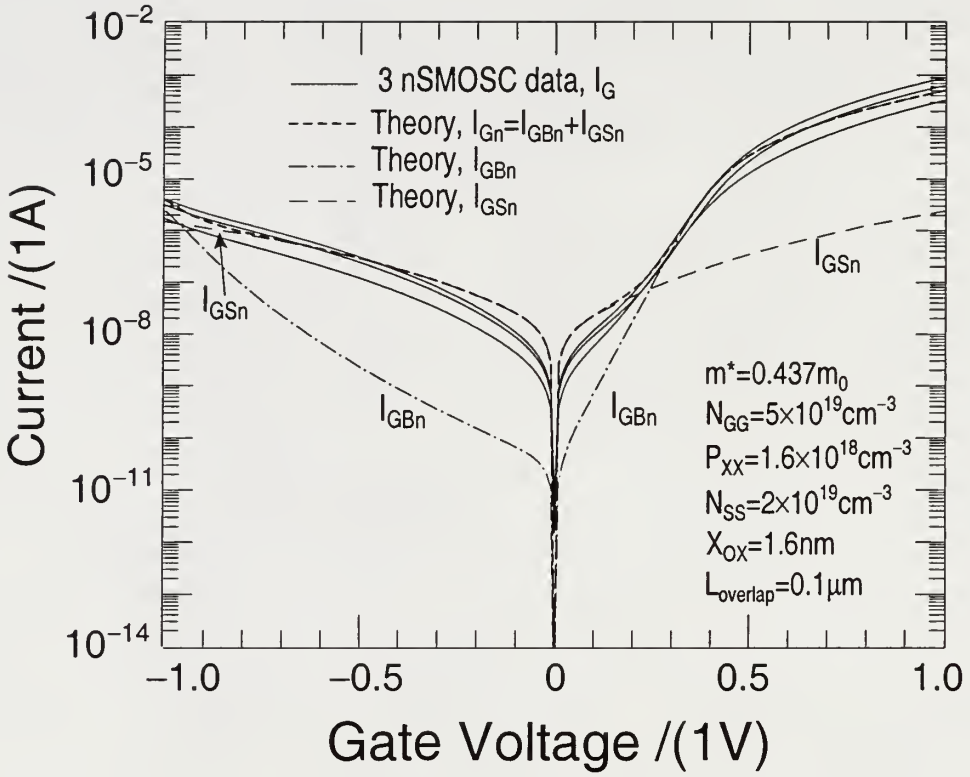

Fig. 4.18 Two geometrical tunneling pathways in n+sourced n+Si/SiO_2/p-Si MOS capacitors. Experimental I_G-V_{GB} curves are matched by adding up two theoretical electron tunneling currents, I_{GBn} for electron tunneling through the channel area, and I_{GSn} for electron tunneling through the G/S overlap area.

experimental correlation of $J_S=I_S/A_{GS}$ in part (b). The same oxide thickness and n+gate doping concentration in part (a) together with two n+ source doping concentrations ($\times 10^{19}cm^{-3}$) of 3.0 and 5.0 were used in the computation of the two theoretical electron tunneling curves in part (b). Noticeably, unlike the case of pSMOSC in which the match is almost perfect, the experimental J_S data in nSMOSC exhibits faster roll-off in the low voltage range than the theory even with the refined WKB tunneling theory. This points to additional differences in electron and hole tunneling which are not considered by the current theory.

In figure 4.18, the experimental gate-terminal current I_G is shown to have two geometric tunneling pathways, one through the channel area described by the $I_{GBn}=J_{GBn}A_{GB}$ theory and the other through the G/S overlap area described by the $I_{GSn}=J_{GSn}A_{GS}$ theory. Similar to fig. 14 for pSMOSCs, this theory-experimental correlation verifies that the inflection point in I_G at 0.25V is from the crossing of two geometric tunneling current component and that the small G/S overlap area is the dominant tunneling pathway in the low voltage range from −1.0V to 0.25V.

The second set of nSMOSC data shown in Figs. 4.19-20 has the following features which are different from those of the previous example: (1) there is no inflection point in the gate current, (2) the ratio of I_B/I_G in the depletion-flatband voltages is about 3 times larger, and (3) the knee point in I_B shifts more than 0.2V in the negative direction. All these changes can be explained by a lower p-well impurity doping concentration of $5\times 10^{17}cm^{-3}$ as compared to $1\text{-}2\times 10^{18}cm^{-3}$ which is demonstrated in the two figures. In figure 4.19, I_G at $V_{GB}>0$ and I_B at $V_{GB}<-0.55V$ are

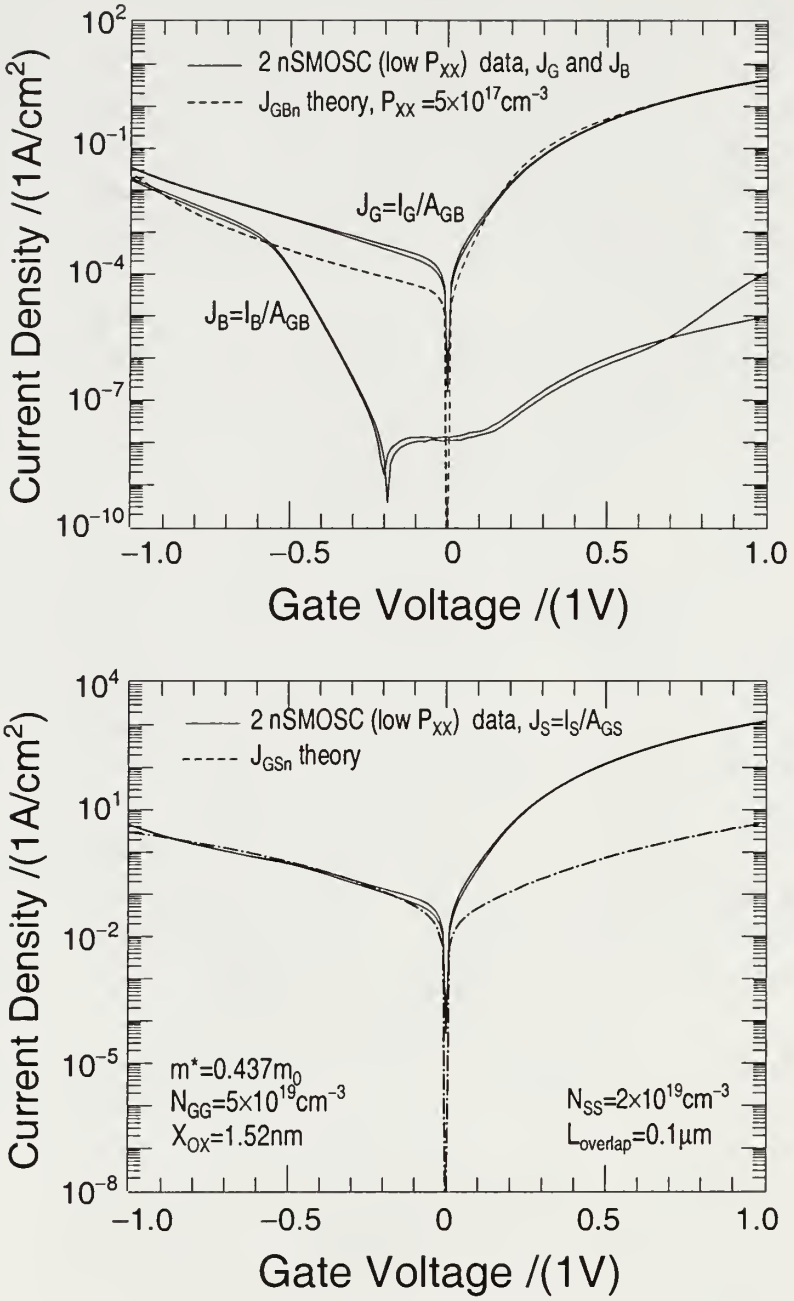

Fig. 4.19 Another theory-experimental correlation for electron tunneling in nSMOSCs with lower p-well doping concentrations than that of the previous three nMOSCs shown in Figs. 17-18. (a) Comparison of the J_{GBn} theory with $J_G=I_G/A_{GB}$ and $J_B=I_B/A_{GB}$. (b) Comparison of the J_{GSn} theory with $J_S=I_S/A_{GS}$. The tunneling theories in the two parts use the same transistor parameters including doping concentrations of $5\times10^{17}cm^{-3}$ in the p-well, $5\times10^{19}cm^{-3}$ in the n+ gate and $2\times10^{19}cm^{-3}$ in the n+ source, and an oxide thickness of 1.52nm.

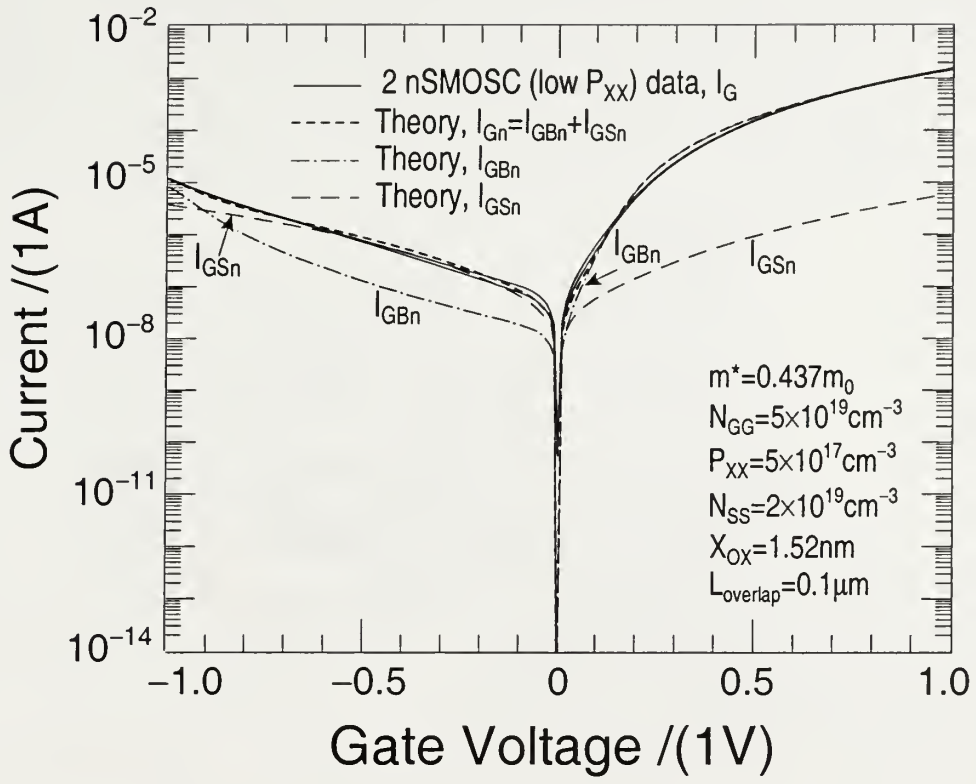

Fig. 4.20 Two geometrical tunneling pathways in the two nSMOSCs with relatively low p-well doping concentrations. Experimental I_G-V_{GB} curves are matched by adding up two theoretical electron tunneling currents, I_{GBn} for electron tunneling through the channel area, and I_{GSn} for electron tunneling through the G/S overlap area.

matched by the theory of electron tunneling through the channel area in part (a), and $I_S \approx I_G$ at $V_{GB}<0$ is matched by the theory of electron tunneling through the G/S overlap area. Transistor parameters that give the best match are: impurity doping concentrations of $5\times10^{17}cm^{-3}$ in the p-well, $5\times10^{19}cm^{-3}$ in the n+ gate and $2\times10^{19}cm^{-3}$ in the n+ source, and oxide thickness of 1.52nm. The lower p-well doping concentration increases the electron tunneling current from n+gate to the p-well (I_B) by reducing the integrated tunneling barrier, refer to figure 4.5(a). The knee point is determined by the threshold voltage for surface (weak) inversion in the p-well which shifts negatively with lower p-well doping concentrations. The lack of knee point in I_G can be readily understood from figure 4.20 in which the two geometric tunneling current components I_{GBn} and I_{GSn} are separated. Due to the increase of the channel component at low positive voltages, I_{GBn} and I_{GSn} no longer crosses at the lower p-well doping concentration. The overlap area component still dominates the gate-terminal current in the negative gate voltage range.

4.6 Summary

Followed by an extensive review of the literatures, we developed a theory that applies to the tunneling of electron, hole and valence electrons through a trapezoidal barrier of the technologically important MOS structure. The new theory refines the traditional WKB tunneling probability by the incorporation of a kinetic energy dependent pre-exponential factor which is important for accurate modeling of MOS tunneling currents at low fields. The theory was applied to study the tunneling currents

from all terminals of thin oxide sourced MOS capacitors in order to identify their specific tunneling carrier species and geometric tunneling pathways. The experimental data on MOS capacitors verifies two conclusions predicted by the theory which also applies to cMOS transistors: (1) hole currents dominate over electron currents in p+gate pMOS transistors, and (2) the gate overlapped source and drain areas pass higher currents than the channel area in the depletion-flatband gate voltage range despite its small area. Application of the tunneling theory to extract transistor design parameters such as doping concentrations and gate oxide thickness was also demonstrated.

CHAPTER 5
INTERBAND TUNNELING NEAR THE DRAIN/SUBSTRATE JUNCTION PERIMETERS IN MOS TRANSISTORS

5.1 Introduction

The miniturization of MOS transistors requires the use of thinner gate oxide and higher surface impurity doping concentration in each new technology generation in order to minimize short channel effects in scaled transistors [102]. As a result, the drain/substrate and source/substrate junction space charge regions become thinner. During transistor operation, they can be further thinned by an accumulation gate voltage which makes the depletion of substrate majority carriers more difficult at the junction perimeter. It is well known that interband p/n junction tunneling could dominate the current-voltage characteristics if the transition region is thinner than about 100Å, as in an Esaki diode [56]. The interband p/n junction tunneling at reverse drain/substrate bias was recognized as a major off-state leakage current path [103]. DCIV [5,8] data on recent generations of MOS transistors show a rising baseline in the accumulation gate voltage range of the I_B-V_{GB} curve, which was leveraged by Sah as a sensitive indicator of surface impurity concentrations [8]. This experimental feature prompted us to look at the theoretical gate voltage V_{GB} dependence of the drain junction perimeter tunneling current I_{PN}, which is presented in section 5.2. We then examine the effect of energy quantization in the surface accumulation layer on I_{PN} and the possibility of probing discrete energy levels from I_{PN}-V_{GB} characteristics in section 5.3. A similar idea was

proposed by Quinn et al. [104] to probe energy levels in the surface inversion layer of a p-type substrate by measuring the tunneling current between a p+source and the n-inversion layer. Detailed calculation of drain/substrate barrier height at the surface reveals that the direct interband tunneling under forward junction bias has a large threshold $|V_{GB}|$, not compatible with experimental DCIV data shown in section 5.4. The I_B-V_{GB} data shows peaked structures superimposed on the rising I_B baseline in the accumulation V_{GB} range. In section 5.5, the I_B peak is correlated to Shockley-Read-Hall (SRH) recombination at interface traps (N_{IT}) located over the gate-overlapped drain region. The evidence of N_{IT} in the overlap region led us to develop a theory for trap assisted tunneling which also has a lower threshold $|V_{GB}|$ to account for the DCIV data, in section 5.6. This tunneling current is similar to the 'excess current' in a bulk Esaki diode, where bulk traps are involved in carrier transport, such as the gold centers in a gold-doped silicon p/n diode studied by Sah [58]. Section 5.7 concludes this chapter.

5.2 Theory of Drain Junction Perimeter Tunneling

Figure 5.1 depicts the energy band diagram of a nMOS transistor with its drain/substrate junction forward-biased at V_{PN}, such as in drain-emitter DCIV measurements. The barrier height V_B of the n+/p junction is increased by a hole accumulation surface potential V_{SX} induced by the gate voltage V_{GB}. If $V_B > E_g/q$, conduction band electrons on the n+D side can tunnel into valence band holes on the p-substrate side, giving rise to the junction perimeter tunneling current in base and drain

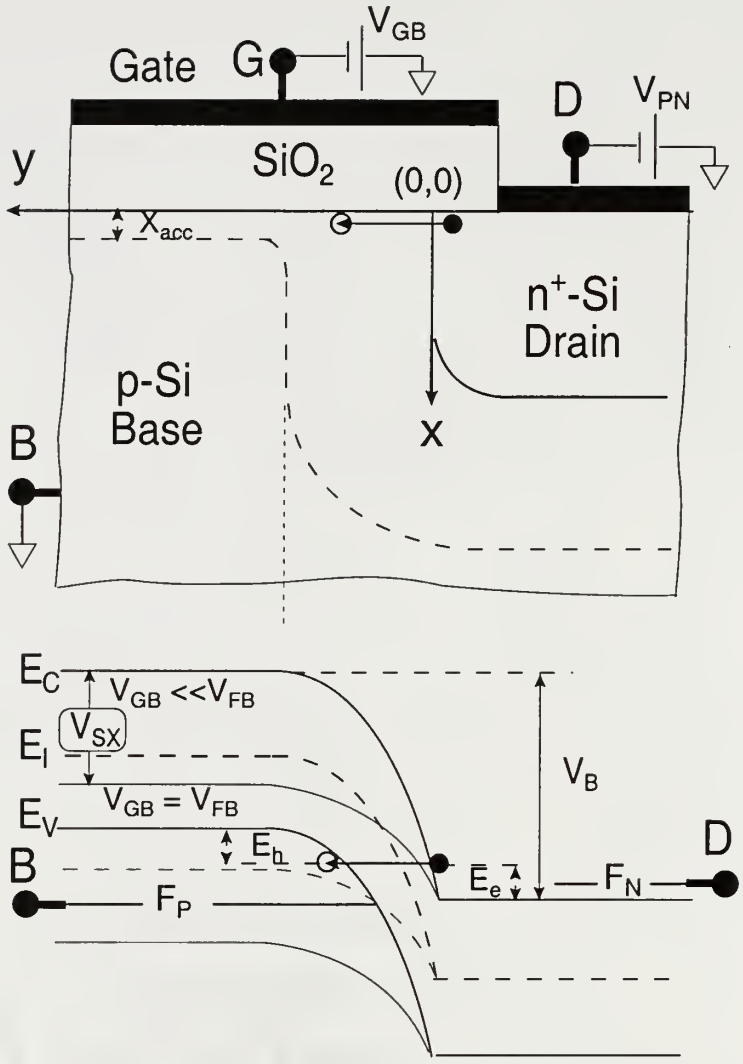

Fig. 5.1 Cross-section and energy band diagram of a n+Drain/p-Base junction of a MOS transistor at X=0 under strong hole accumulation gate bias. Electrons at the n+ side with kinetic energy E_e tunnel into hole states with kinetic energy E_h. V_B = n+/p barrier height, V_{BI}= built-in potential of the n+/p junction, V_{SX} = surface band bending in the hole accumulation layer with a thickness of x_{acc}, and V_{PN} = forward bias voltage applied to the n+/p junction.

terminals. Energy conservation requires that the electron kinetic energy of the initial state E_e and hole kinetic energy of the final state E_h satisfy the following equation

$$E_e + E_h = qV_B - E_g \tag{5.1}$$

The general approach used in chapter 4 for the derivation of gate oxide tunneling theory also applies to the interband p/n junction tunneling at the surface with distinctions in (1) tunneling effective mass and (2) shape of the tunneling barrier. Since the initial electron state is in the silicon conduction band near k=0.85(2π/a), where a=5.43Å is the lattice constant of silicon crystal, while the final hole state is in the valence band near k=0, strict conservation of transverse (perpendicular to the tunneling direction) momentum k_t must involve phonons. The theory of phonon-assisted tunneling was developed by Kledysh [105] and Price and Radcliffe [106], and was applied to p/n tunnel junction by Kane [57]. The theory is rather laborious, and only valid for constant electric field in the barrier region. In our case of p/n junction at the surface, interface roughness scattering provides another means of conserving transverse momentum. To derive the essential features, we shall relax the requirement of k_t conservation. The resultant formula for interband tunneling with the initial state in one of the six silicon conduction band valleys is [107]:

$$J_{PN} = \frac{4\pi\eta q m_d}{h^3} \int_0^{qV_B-E_g} T(E)\,[f(E)-f(E+qV_{PN})]\,E\,dE \tag{5.2}$$

The integration is over electron kinetic energy E that have a direct tunneling pathway. f(E) is the Fermi function that describes electron occupation of the initial state at the n+ side while $f(E+qV_{PN})$ is the electron occupation factor of the final state at the p side. η

and m_d are respectively the degeneracy factor and the density-of-states effective mass of the subvalley. Assume x, y and z are all along <100> directions, which is consistent with MOS transistor manufacturing practice, the four transverse valleys would give the most contribution to tunneling because of their lower effective mass in the tunneling direction (y). Therefore $\eta=4$ and $m_d=(m_x m_z)^{1/2} = (m_t m_l)^{1/2} = (0.916\times0.191)^{1/2}m_0 = 0.418m_0$. T(E) is the barrier transmission probability which has the following expression from WKB approximation:

$$T(E) \approx \exp\left[-2\int_{Y_1}^{Y_2} \kappa(y)\,dy\right] \qquad (5.3)$$

κ is the imaginary part of the wave number in the energy gap of silicon, Y is the tunneling direction, and y_1 and y_2 are the boundaries of the classical forbidden region. There is no potential discontinuity at the two boundaries, thus the unity pre-exponential factor is a good approximation. Notice that in the forbidden region, in addition to the decaying wave in y direction, there are also the propagating waves in the perpendicular x and z directions.

The simplest description of κ is a parabolic $E(\kappa)$ relationship. A more rigorous form of of κ can be obtained by analytical continuation of the E(k) relationship in the energy band into the energy gap. The result is the a non-parabolic formula which reduces to parabolic relationship at both conduction band edge and valence band edge, which was used by Kane [108] for InSb, and also by Maserjian [109] for SiO_2. The two alternative $E(\kappa)$ relationships are:

$$\hbar^2\kappa^2-\hbar^2k_x^2-\hbar^2k_z^2 = 2m^* \cdot [qV(y)-E] \qquad \text{(parabolic)} \qquad (5.4)$$

$$= 2m^* \cdot [qV(y)-E] \cdot \{1-[qV(y)-E]/E_g\} \qquad \text{(non-parabolic)} \qquad (5.5)$$

m^* is the tunneling effective mass, $m^* = 2\, m_c m_v/(m_c+m_v)$ [107], where $m_c=m_t=0.191m_0$ is the normal mass of the four transverse subvalleys, and, m_v, the light hole mass near $\mathbf{k}=0$ is $m_0/(A+B) = m_0/(4.28+0.75) = 0.199m_0$ [110]. The result is $m^*=0.195m_0$. V(y) is the electro-static potential in the gap region, referenced to the conduction band edge at the n+ side. Eg is the indirect energy gap of 1.12eV in Si. In the non-parabolic formula, $\kappa=0$ defines the two end points y_1 and y_2 of the tunnel region. It can be seen that for non-zero k_x and k_z, the tunneling distance and effective barrier height are both increased, which results in exponential drop of tunneling probability [57,111]. For simplicity, we only consider initial states with nearly zero transverse momentum that has the largest tunneling probability, i.e., $k_x=k_z=0$ in (5.4) and (5.5).

For a parabolic potential barrier (or linear field), only the parabolic formula (5.4) can be integrated. Under depletion approximation and $|E_y|>>|E_x|$ at the surface for the one-sided n+/p junction, the potential on the p-side is:

$$V(y) = (qN_{AA}/2\varepsilon_{si})\ [d^2-(d-y)^2] \qquad (5.6)$$

N_{AA} is the doping concentration at the surface of p-substrate, and d is the depletion layer thickness in y. Using (5.4), we derive

$$T(E) = \exp\{-\sqrt{(2m^*/\hbar^2)\ (2\varepsilon_{si}/q^2N_{AA})}\ [(qV_B-E)(\pi/2 - \sin^{-1}\sqrt{(qV_B-E-E_g)/(qV_B-E)}] - \sqrt{E_g(qV_B-E-E_g)}]\} \qquad \text{(linear F, parabolic)} \qquad (5.7)$$

V_B is the barrier height controlled by V_{GB}.

For the simpler constant field case, i.e., V=qFx, where F denotes the constant field, T is not a function of E since tunneling electrons at different energy levels experience the same triangular barrier. Analytical results for T are obtained for both E(κ) relationships:

$$T(E) = T = \exp[-\alpha\sqrt{m^*}E_g^{3/2}/(q\hbar F)]$$

$$\alpha = 4\sqrt{2}/3 = 1.886 \quad \text{(constant F, parabolic)} \qquad (5.8)$$

$$\alpha = \pi/(2\sqrt{2}) = 1.111 \quad \text{(constant F, non-parabolic)} \qquad (5.9)$$

If the same effective mass is used in (5.8) and (5.9), the non-parabolic formula (5.9) gives an effectively lower tunneling barrier and higher tunneling probability.

As a first order approximation, barriers of any shape can be approximated by a linear barrier with a slope equal to the average field F_{avg}. For a parabolic barrier described by (5.6), the average field in the tunnel region ($y_1<y<y_2$) is:

$$F_{avg} = [F(y_1)+F(y_2)]/2$$

$$= \sqrt{N_{AA}/2\varepsilon_{si}}\ (\sqrt{qV_B} + \sqrt{qV_B-E_g}) \qquad (5.10)$$

Figure 5.2 gives the theoretical J_{PN}-V_{GB} dependence for a gated n+/p junction using non-parabolic formula (5.5) with triangular barrier approximation (5.10) in part (a), and parabolic formula (5.4) with the more exact parabolic barrier (5.6) in part (b). The tunneling effective mass in part (b) is scaled to that of part (a) according to eqs. (5.8) and (5.9) by a factor of $(1.111/1.886)^2$=0.347. The parabolic barrier gives lower tunneling current density than the triangular barrier, due to the T(E) dependence in (5.7), which drops sharply with increasing electron kinetic energy. At an oxide thickness of 1.5nm, under large negative V_{GB}, the interband tunneling current J_{PN} is about 4-5 orders

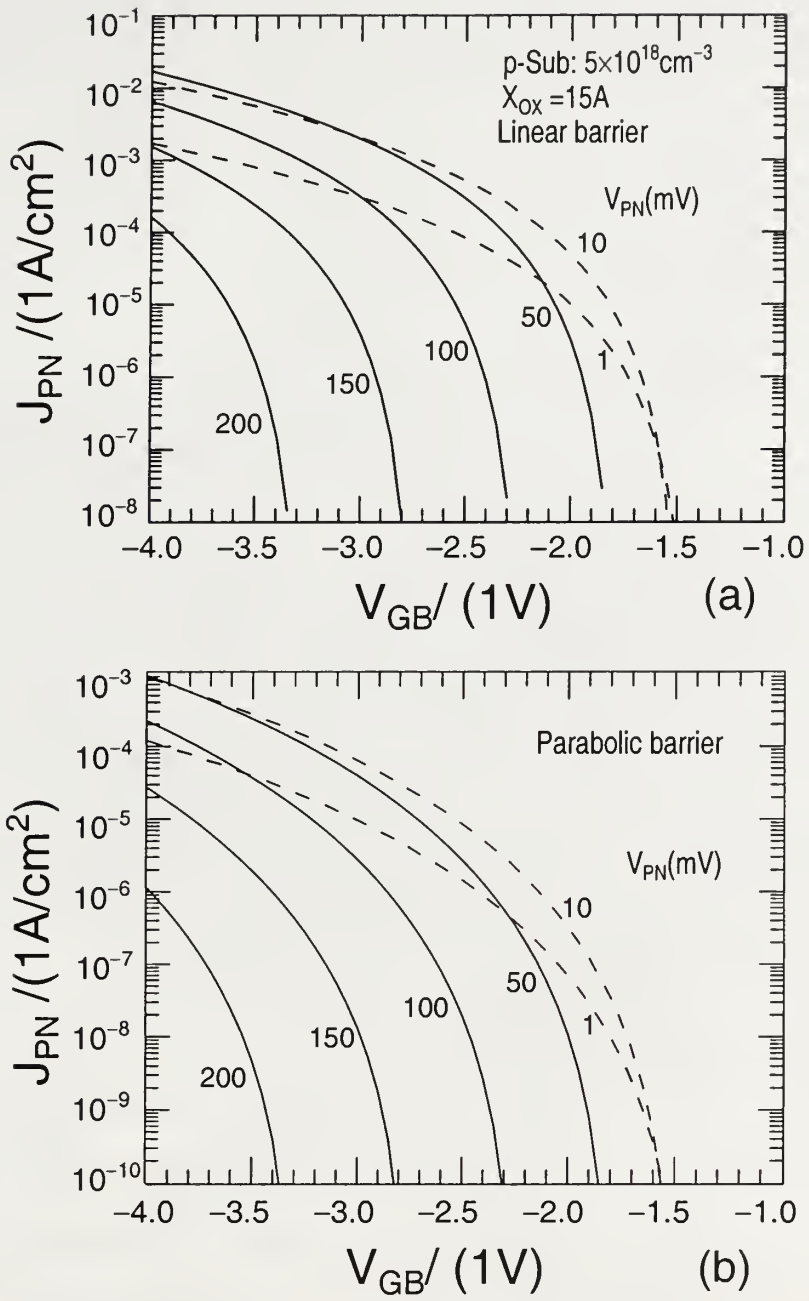

Fig. 5.2 Gate voltage dependence of surface p/n junction perimeter tunneling current density, calculated from two theories: **(a)** average field approximation with non-parabolic E(κ) relationship **(b)** linear field (parabolic potential barrier) with parabolic E(κ) relationship. Junction forward bias changes from 1mV to 200mV as labeled. $N_{AA} = 5\times10^{18}$ cm^{-3} in p-base, $N_{DD} = 5\times10^{19}$ cm^{-3} in n+Gate and gate oxide thickness $X_{OX} = 1.5$nm.

of magnitude smaller than the gate oxide tunneling current J_{OX} [see Fig. 4.8(b)], for a p-substrate doping concentration of $5\times10^{18}cm^{-3}$. The threshold $|V_{GB}|$ increases from 1.5V to 3.3V as the junction forward bias is raised from 1mV to 200mV. Due to this threshold voltage change, at a fixed V_{GB}, J_{PN} decreases as V_{PN} increases except for the two dashed lines at V_{PN}=1mV and 10mV. The drop of J_{PN} at V_{PN}<10mV is due to increased electron tunneling in the reverse direction from p-sub side to n+D side. Two more theoretical figures are given which were computed from the non-parabolic formula.

Figure 5.3 shows the oxide thickness (X_{OX}) dependence of J_{PN}-V_{GB} characteristics. The threshold $|V_{GB}|$ increases almost linearly with X_{OX}, which is expected from: $\Delta V_{GB} = \Delta V_{OX}$ in the strong accumulation gate voltage range. At all oxide thickness, J_{PN} is insignificant as compared to J_{OX} and thus its experimental detection will be masked by the gate oxide tunneling current. J_{PN}, however, is a strong function of p-substrate doping concentration (N_{AA}) while J_{OX} is not. Figure 5.4 shows that J_{PN} would overtake J_{OX} at $N_{AA}>10^{19}cm^{-3}$, for X_{OX}=2.0nm and V_{PN}=50mV. J_{OX} is calculated using the trapezoidal SiO_2 tunneling formula (4.3), which is nearly independent of N_{AA} at $-4V<V_{GB}<-2V$.

5.3 Tunneling into Quantized Surface Accumulation Layer

The proceeding J_{PN} formulas assumed that both the initial state and final state of the tunneling transition can be described by three-dimensional (3D) states characterized by crystal momentum (k_x,k_y,k_z). It was recognized in 1957 by Schrieffer [112] that

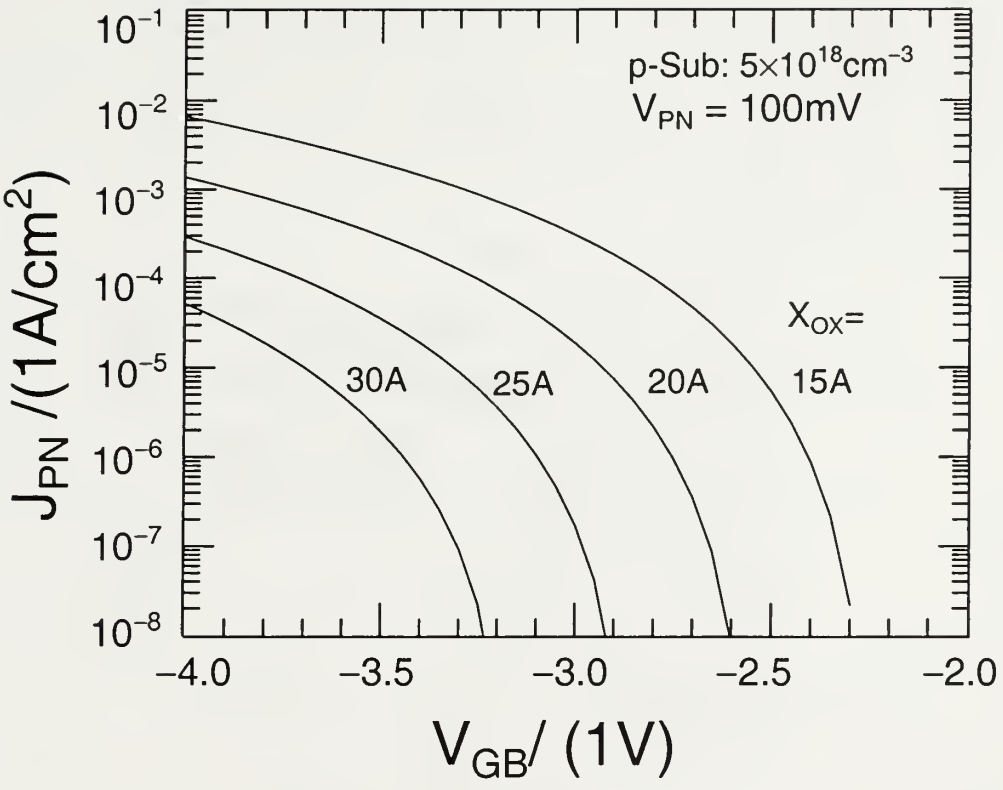

Fig. 5.3 Theoretical surface p/n junction tunneling current density as a function of gate voltage, for various oxide thickness from 1.5nm to 3.0nm. Forward bias V_{PN}=100mV, p-Base and gate doping concentrations are the same as those in Fig. 5.2.

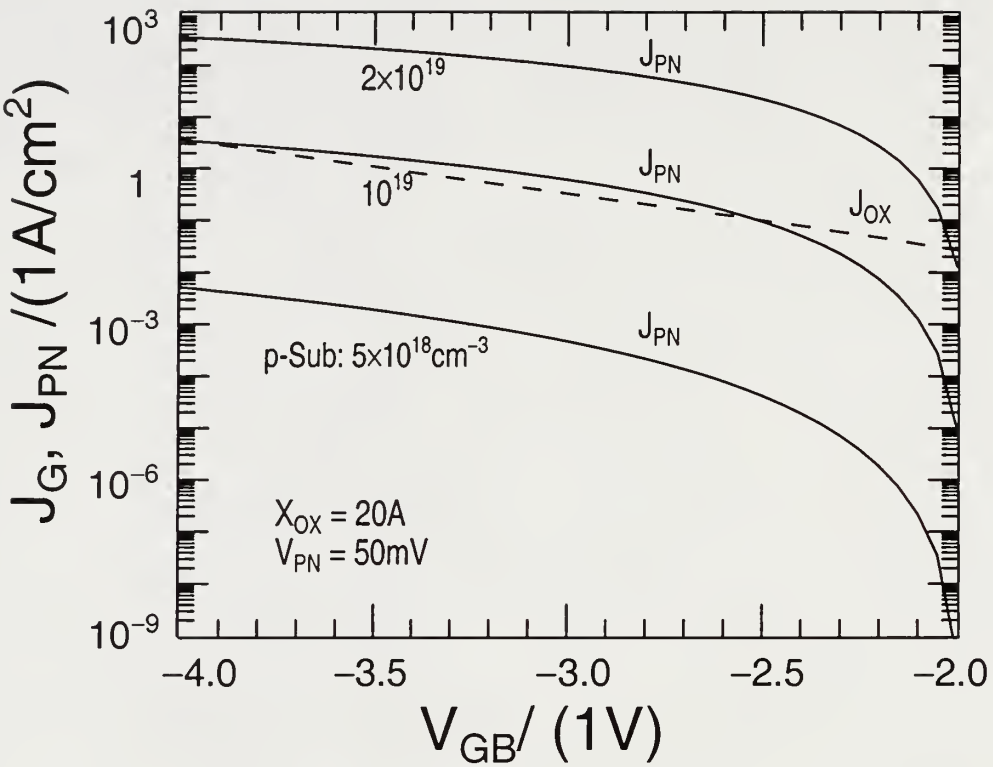

Fig. 5.4 Comparison of surface p/n junction tunneling current density J_{PN} (solid lines) and gate oxide tunneling current density J_{OX} (dashed line), for three p-Base doping concentrations. X_{OX} = 2.0nm and V_{PN} = 50mV.

strong space-charge field due to a large bias (V_{GB}) would confine the motion of carriers in a thin surface layer and quantize the energy levels in the direction perpendicular to the surface. This quantization effect has been studied for both surface inversion layers [98,113] and surface accumulation layers [114] in MOS structures. In this section, we will develop a theory for surface electron tunneling from 3D states in n+drain to 2D states in the hole accumulation layer of p-substrate, and evaluate the effect of quantized energy levels on J_{PN}-V_{GB} characteristics.

In the quantized hole accumulation layer, hole kinetic energy is described by:

$$E_h = E_n + (\hbar^2/2m_v)(k_y^2 + k_z^2) \qquad (5.11)$$

$$= qV_B - E_g - E$$

where E_n denotes the discrete energy level in the x-direction, and E is the electron kinetic energy in the initial 3D state. The density-of-states in the 2D hole accumulation layer is:

$$D_2(E_h) = \frac{2A}{(2\pi)^2} \frac{2\pi k\, dk}{dE_h} S(E_n) = \frac{A\, m_v}{\pi\hbar^2} S(E_h) \qquad (5.12)$$

where $S(E_h) = 0, \quad 0 < E_h < E_0$

$$= n, \quad E_{n-1} < E_h < E_n \quad (n=1, 2\ldots)$$

A is the area of the 2D layer, E_0 is the energy level of the ground state, and E_n is the energy level of the n-th excited state. $S(E_h)$ is a staircase function with increasing step of unity at E_h=E_n. The usual 3D density of hole states neglecting quantization effect in x-direction is:

$$D_3(E_h) = \frac{2V}{(2\pi)^3} \frac{4\pi k^2 dk}{dE_h} = \frac{V\sqrt{2m_v^3 E_h}}{\pi^2 \hbar^3} \qquad (5.13)$$

V si the volume of the 3D system and $V/A = X_{acc}$, a characteristic thickness of the surface accumulation layer.

Since the tunneling probability is directly proportional to the density of states of the final state, we derive a J_{PN} formula for tunneling into 2D states from (5.2), (5.11), (5.12) and (5.13):

$$J_{PN} = \frac{4\pi\eta q\sqrt{m_x m_z}}{h^3} \int_0^{qV_B - E_g} T(E)[f(E) - f(E+qV_{PN})]E \times \left[\frac{\pi\hbar/x_{acc}}{\sqrt{2m_v(qV_B - E_g - E)}} S(E_h)\right] dE \qquad (5.14)$$

The additional factor in [ ] equals $D_2(E_h)/D_3(E_h)$. The staircase function $S(E_h)$ in (5.14) predicts break points in J_{PN}-V_{GB} curves, which corresponds to the crossing of n+ drain conduction band minimum and discrete hole energy levels on the p-side.

We will next use the simple triangular well approximation to estimate the discrete energy levels in the surface accumulation layer. Stern showed that this simple approximation gives good analytical results for surface inversion layers [98]. However, using the surface electric field F_S as the slope of the triangular well would give a ground state level E_0 greater than the depth of the well given by the surface band bending in the accumulation layer, which is not self-consistent. This is because the accumulation layer has a much smaller surface band bending than the inversion layer due to the lack of a depletion layer. We will use a slope which is scaled down from F_S. A scaling factor of

0.03 would get the ground state energy level in the range of those obtained from more rigorous numerical solutions of Ando [114] and Sune et al. [115]. This gives the following expression [98] for the energy levels:

$$E_n = (\hbar^2/2m_v)^{1/3} \cdot [(3\pi q/2)(0.03F_s)(n+3/4)]^{2/3} \quad (n=0,1,2...) \quad (5.15)$$

The accumulation layer thickness can be estimated from the x-distance of ground state holes from the Si/SiO_2 interface (X=0) as follows [98]:

$$X_{acc} = 2E_0/[3q(0.03F_s)] \quad (5.16)$$

Figure 5.5 gives an example of the 2D density of hole states normalized by $(m_v/\pi\hbar^2)$ in the surface accumulation layer at $F_S=10^6$V/cm. Figure 5.6 shows gate voltage dependence of the lowest four bound state energy levels (solid lines) and the surface band bending qV_{SX} (dashed line). V_{SX} was calculated using Fermi-Dirac statistics in gate and drain layers. Since $q|V_{SX}|$ indicates the depth of the surface potential well, the four energy levels are all below $q|V_{SX}|$ and therefore are consistent with one-dimensional bound states.

We will next give a theoretical illustration of quantization effects on surface p/n junction tunneling characteristics. J_{PN} and its derivative dJ_{PN}/dV_{GB} are plotted in Fig. 5.7(a) and Fig. 5.7(b) respectively. The solid lines were computed from (5.14) using the 2D DOS function of (5.12), with $S_1=1$, $S_2=2$ and $S_3=3$. The dashed lines are from two hypothetical 2D DOS functions: $S_1=S_2=S_3=1$ for the long-dash lines and $S_0=1$, $S_1=2$, $S_3=2$ for the short-dash lines. Two breakpoints at $V_{GB}=-2.79$V and -3.73V in the solid line can be readily observed from the derivative plot in Fig. 5.7(b), also visible in Fig. 5.7(a) with the help of the dash lines. These break points correspond to the crossing

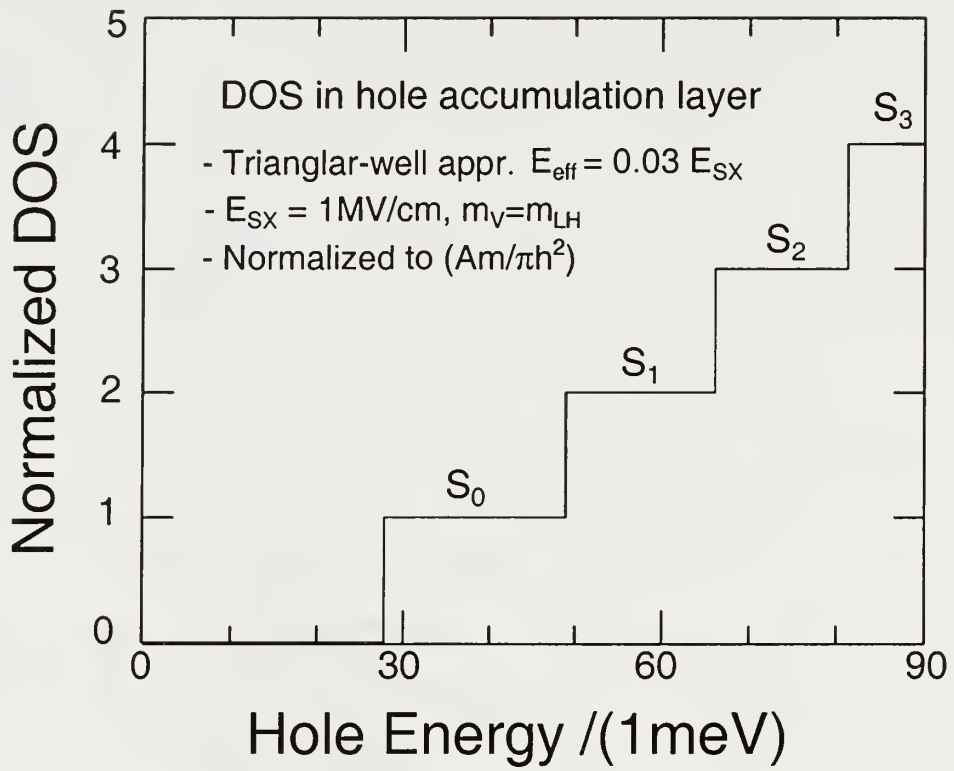

Fig. 5.5 Normalized density of hole states in a quantized suface accumulation layer with a surface electric field E_{SX} = 1MV/cm. Quantum leaps occur when hole energy E_h= nth energy level of one-dimensional hole bound states. The light hole band normal effective mass $m_V = 0.199m_0$ was used.

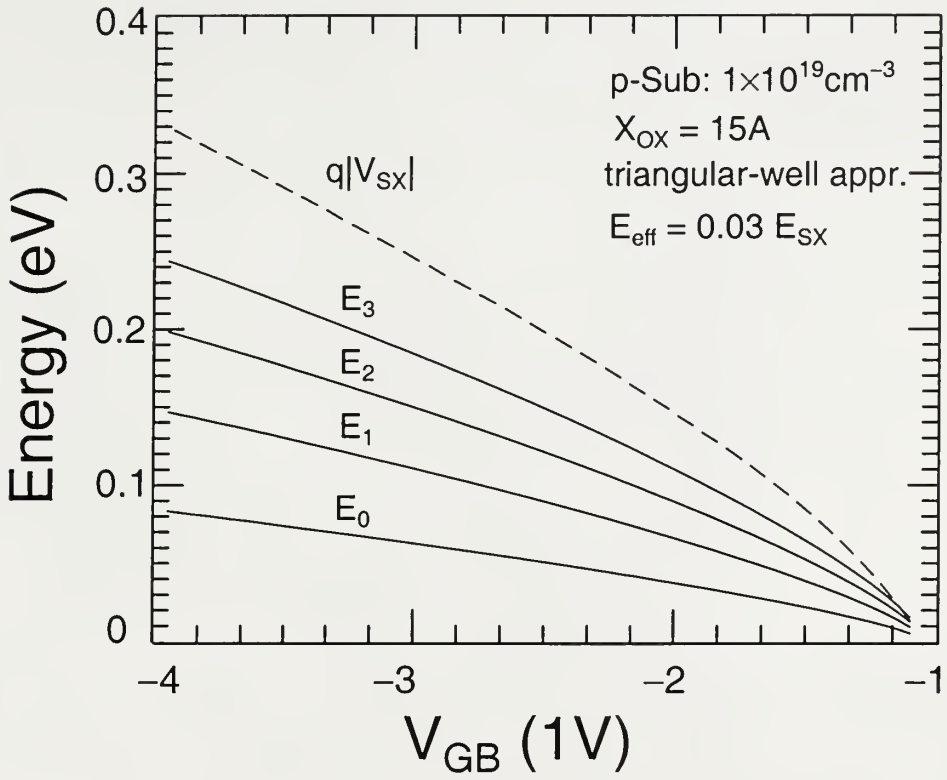

Fig. 5.6 Gate voltage dependence of the four lowest energy levels of one-dimensional hole bound states in a surface accumulation layer with the band bending shown as the dash line. Same theoretical model as in Fog. 5.5 for the discrete energy levels. $N_{AA}=10^{19}cm^{-3}$ in p-base and $X_{OX}=1.5nm$.

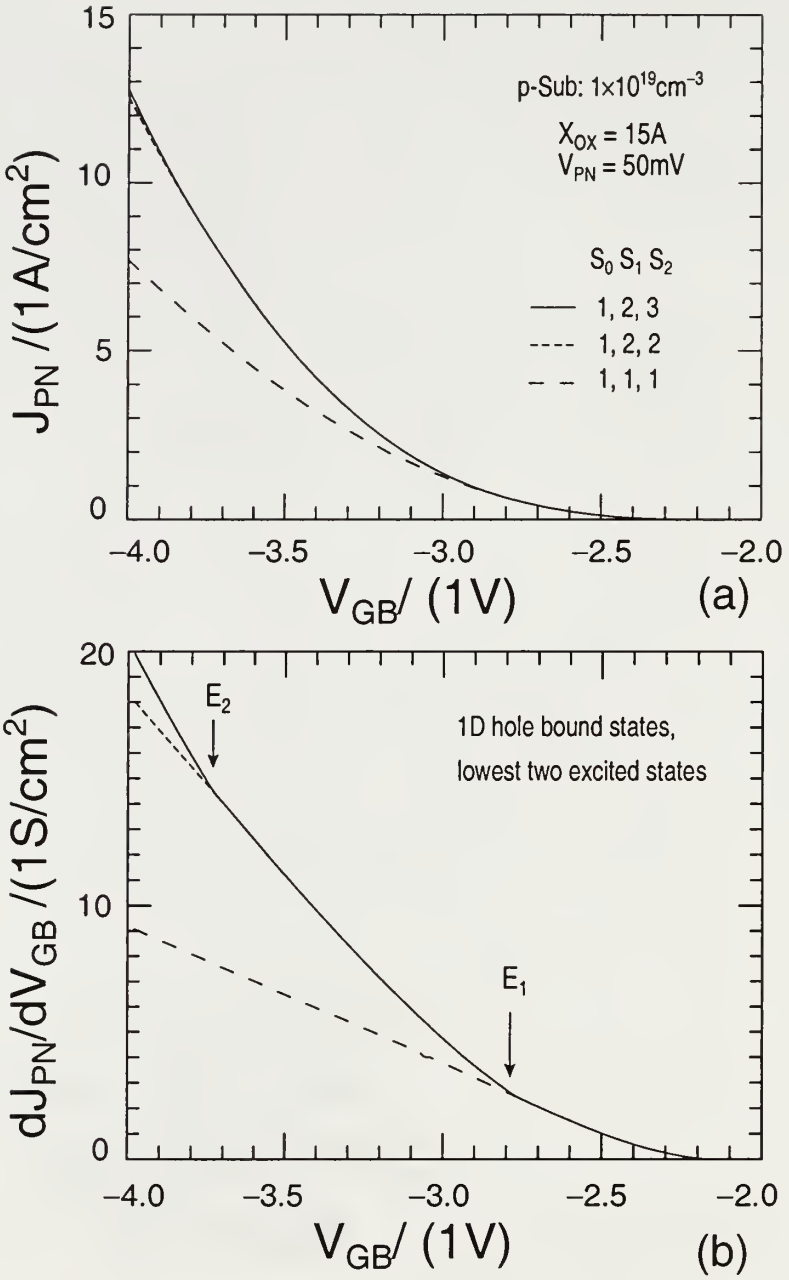

Fig. 5.7 Effects of quantized hole accumulation layer on surface p/n junction tunneling current. **(a)** Tunneling current- gate voltage characteristcs and **(b)** its first derivative. The solid lines were computed from theoretical staircase 2D density of state in (5.12), while the dashed line used two hypothetical DOS functions. The two breakpoints due to the discontinuity of DOS function at $E_h=E_1$ and $E_h=E_2$ are more obvious in part (b). $N_{AA}=10^{19}cm^{-3}$ in p-base, $X_{OX}=1.5nm$ and $V_{PN}=50mV$.

of the n+ side conduction band edge with the first excited and second excited hole state energy levels on the p-substrate side respectively.

We conclude this section by noting that direct surface p/n junction tunneling is not significant in current MOS technologies with substrate doping concentrations on the order of 10^{17}-$10^{18}cm^{-3}$. In samples with high substrate doping concentration so that J_{PN} dominates J_{OX}, it may be possible to gauge quantized energy levels in the surface accumulation layer from the breaks in the J_{PN}-V_{GB} characteristics. High-order derivatives may be needed to enhance the sensitivity to detect the quantum effects under finite temperatures since the one-dimensional discrete energy levels would be broadened to yield a two-dimensional DOS with no sharp steps.

5.4 Base-Terminal Current Under Accumulation Gate Voltages

Recent experimental DCIV data on thin oxide (< 5nm) MOS transistors [5,116] show base-terminal current I_B that rises with accumulation gate voltage $|V_{GB}|$. Figure 5.8 gives examples of three MOST's from two thin gate oxide technologies: a 4.5nm technology in part (a) and a 3.0nm technology in part (b). To focus on the accumulation V_{GB} range, in part (b), we only show nMOST data at $V_{GB}<0$ and pMOST data at $V_{GB}>0$. The Families of DCIV curves correspond to top-emitter forward biases from 50mV to 350mV on a W/L= 20μm/0.25μm pMOST in part (a) and drain-emitter forward biases from 200mV to 400mV on W/L=10μm/0.32μm MOSTs in part (b). All I_B-V_{GB} curves show peaked structures on top of the baselines that rise with accumulation $|V_{GB}|$.

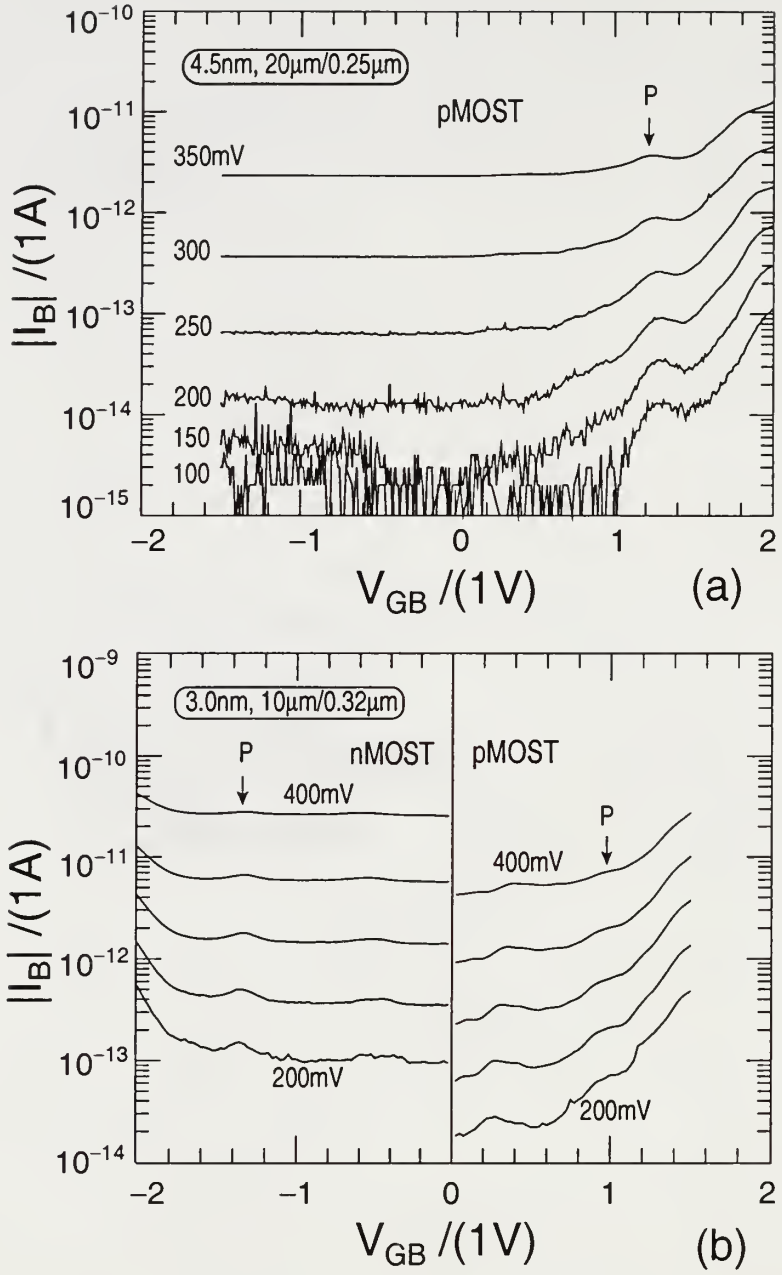

Fig. 5.8 Experimental DCIV data with rising I_B in the accumulation gate voltage range. **(a)** TE-DCIV data from a 4.5nm oxide technology, W/L=20µ/0.25µm, pMOST. The I_B-V_{GB} curves correspond to TE forward bias of 100mV to 350mV. **(b)** DE-DCIV data from a 3.0nm oxide technology, W/L=10µm/0.32µm for both nMOST (left) and pMOST (right). DE forward bias varies from 200mV to 400mV 200mV to 400mV with a step of 50mV. The arrows point to I_B peaks in the accumulation gate voltage range.

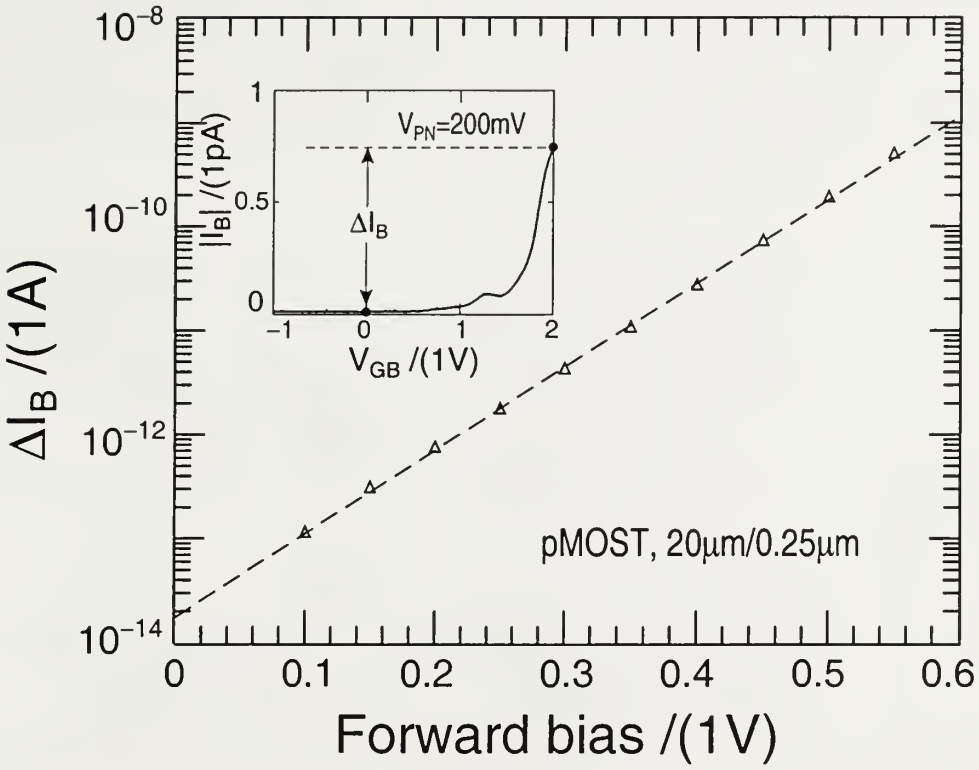

Fig. 5.9 Forward bias V_{PN} dependence of the difference base-terminal current ΔI_B for the 20μm/0.25μm pMOST in Fig. 5.8(a). $\Delta I_B = |I_B(2V) - I_B(0V)|$, see inset. The dash line is a least-squares-fit to $\Delta I_B = I_0 \exp(qV_{PN}/nkT)$ which gives n=2.11±0.02.

The rising baselines are mostly from currents flowing through the emitter/base p/n junctions, which were verified by comparing the emitter current with the base current. The following features of the baselines are not consistent with the theories of surface p/n junction tunneling: (1) Experimental $I_{B\text{-}acc}$ increases with V_{PN} as well as $|V_{GB}|$ while theoretical J_{PN} decreases with V_{PN} (see Fig. 5.2); (2) The rising baselines have low threshold $|V_{GB}|$ around 1V which is almost independent of V_{PN} while the J_{PN} theory predicts a significant increase of the threshold at higher V_{PN}. In particular, we analyzed the V_{PN} dependence of I_B at V_{GB}=2V for the 20μm/0.25μm pMOST in Fig. 5.9. The least-squares-fit (dash line) of $\Delta I_B = |I_B(2V) - I_B(0V)|$ to $\exp(qV_{PN}/nkT)$ determines an ideality factor of n=2.11±0.02.

The theory-experimental discrepancy can be understood from the finite doping concentrations in drain and source regions. The ideal J_{PN} theory neglects surface band bendings in the gate-overlapped emitter region, i.e., V_{SE}=0, which is valid in the limit of infinite doping concentrations. In real MOSTs as shown by Fig. 2.5, under channel accumulation gate voltages, the energy bands in the overlap region bend towards carrier depletion and even inversion conditions, in the same direction of the energy bend bending in the channel region, thus decreasing the potential barrier across emitter/base p/n junction at the surface which prevents the direct p/n junction tunneling. In Fig. 5.10(a), we compare the x-direction surface band bending in the overlap region (V_{SE}) for five n+ drain doping concentrations (N_{DD}) from $10^{18}cm^{-3}$ to $10^{20}cm^{-3}$. The dash line gives the x-direction surface band bending in the p-substrate (V_{SX}) with a doping concentration of $10^{17}cm^{-3}$. The potential barrier height of the p/n junction at the surface

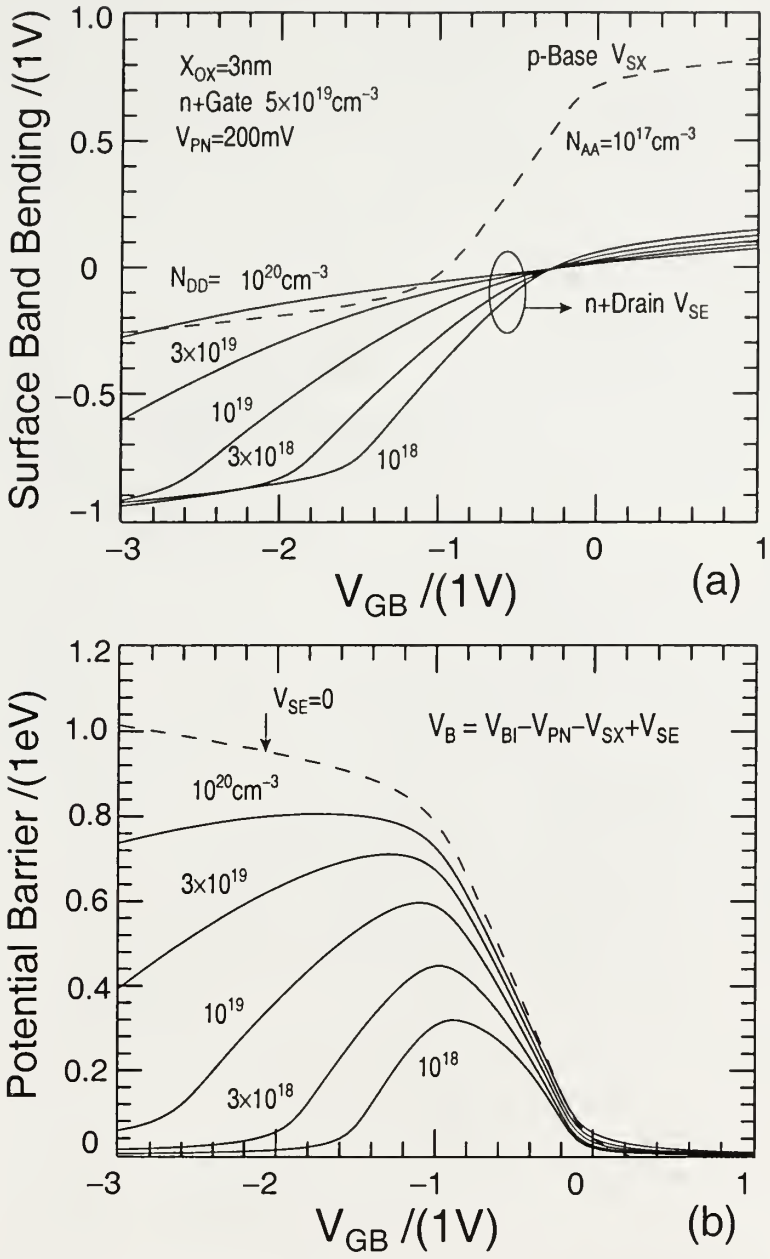

Fig. 5.10 Effects of finite n+ drain doping concentrations on **(a)** surface band bending in gate-overlapped drain region and **(b)** potential barrier height at the surface (X=0) of the n+drain/p-base junction. The five solid lines correspond to five n+ drain doping concentrations. The dash line in part (a) is the surface band bending in the p-base and another dash line in part (b) shows the potential barrier for infinitely large doping concentrations in n+ drain. Parameters used include: $N_{AA}=10^{17}cm^{-3}$ in p-base, $5\times10^{19}cm^{-3}$ in n+ gate and X_{OX}=3.0nm.

is then computed from eq. (2.18) and shown in Fig. 5.10(b). As the drain doping concentration increases from $10^{18}cm^{-3}$ to $10^{20}cm^{-3}$, the maximum barrier height increases from 0.3eV to about 0.8eV under a junction forward bias of 0.2V. In all cases, the barrier height drops at large $|V_{GB}|$ due to carrier depletion at the drain surface, in contrast to the ideal case (dash line) with V_{SE}=0 that keeps rising at large $|V_{GB}|$. Fig. 5.10(b) clearly shows that the potential barrier height at the surface of the drain/base junction does not exceed E_g=1.12eV which prevents direct surface p/n junction tunneling in MOS transistors. In general, interband tunneling in silicon requires a minimum potential barrier of E_g=1.12eV to cross the conduction band and valence band. However, if there are gap states such as those from interface traps, the minimum barrier height would be lowered for tunneling through the gap states [58].

DCIV curves in Figs. 5.8(a) and 5.8(b) all show peaked structures as pointed to by arrows on top of $I_{B\text{-}acc}$, most prominent in the 20μm/0.25μm pMOST. The peaks are around −1.4V for the nMOST and between −1.0V to −1.3V for the pMOSTs. These peak gate voltages and their shift with V_{PN} (positive for nMOST and negative for pMOST) are characteristic of SRH recombination at interface traps in the LDER overlapped by the gate (see Figs. 3.3 and 3.7 in chapter 2). Smaller peaks not labeled are visible in Fig. 5.8(b) at the weak inversion $|V_{GB}|$ around 0.3V-0.4V, which are from SRH recombination at N_{IT}'s located over the MCR.

The existence of interface traps in the gate-overlapped LDER suggests interface trap assisted tunneling as the origin for the rising $I_{B\text{-}acc}$, which will be analyzed in detail in the next section. In chapter 3, we demonstrated the extraction of transistor parameters

such as channel doping concentrations from the DCIV peak around the weak inversion V_{GB} range. We will next show a similar application to determine the doping concentration in the LDER from the DCIV peak in the accumulation V_{GB} range.

Figure 5.11 compares theoretical SRH recombination rate R_{SS1} (dash lines) in the LDER with the experimental DCIV peak in the 20μm/0.25μm pMOST (solid lines). Theoretical SRH recombination rate per unit interface trap was obtained using a mid-gap interface trap energy level E_{TI}=0 and equal electron and hole capture rate coefficients c_{ns}=c_{ps}=$10^{-8}cm^3s^{-1}$. The nominal oxide thickness of 4.5nm for the technology was used in the theory. The best theory-experiment match in peak gate voltage as shown in Fig. 5.11, was achieved with N_{AA}=$2.0\times10^{18}cm^{-3}$ in the LDER of the pMOST.

The variation of peak gate voltage with LDER impurity doping concentration is shown in Fig. 5.12. Under a p+drain/n-base junction forward bias of 100mV, the SRH recombination rate peak ($R_{SS1\text{-}pk}$) shifts from 0.9V to 2.3V as the LDER doping concentration increases from $5\times10^{17}cm^{-3}$ to $1.5\times10^{19}cm^{-3}$. This demonstrates the sensitivity of the technique which is better than 10% for the typical mid-$10^{18}cm^{-3}$ range provided that the experimental peak voltage can be determined within 50mV. The I_B peak in the pMOST data of Fig. 5.8(a) was selected for the fit in Fig. 5.11. For the two MOST's in Fig. 5.8(b), the theory in Fig. 5.12 predicts LDER doping concentrations of $2.5\times10^{18}cm^{-3}$ in the nMOST from $V_{GB\text{-}pk}$=1.40V and $7\times10^{17}cm^{-3}$ in the pMOST from $V_{GB\text{-}pk}$=0.95V.

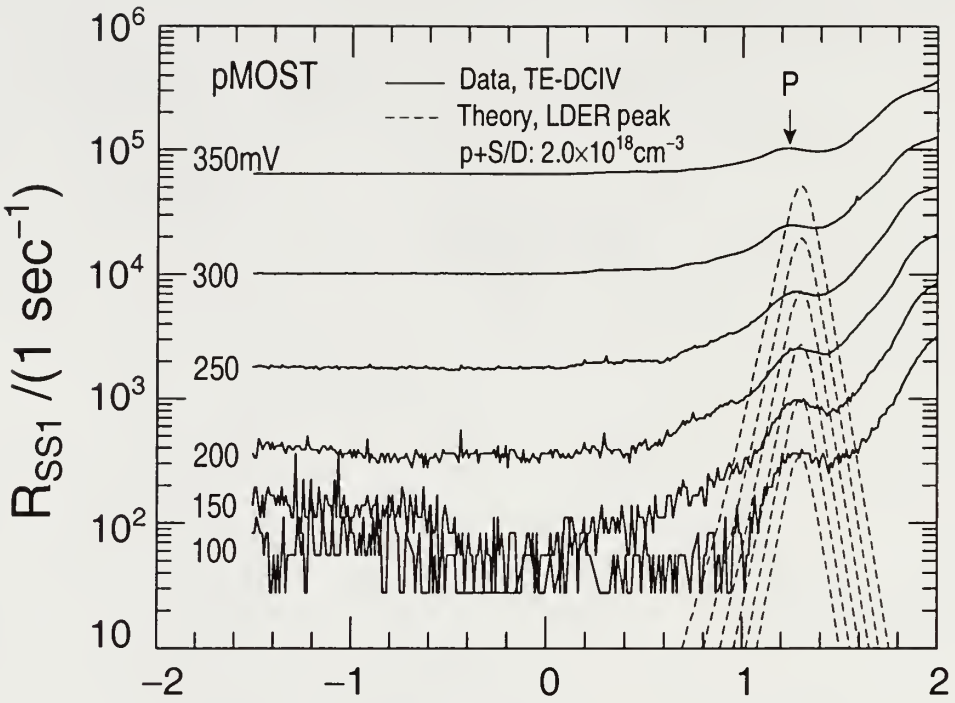

Fig. 5.11 Theory-experiment comparison of I_B peak in the accumulation gate voltage range. Experimental TE-DCIV data are those of the 20μm/0.25μm pMOST presented in Fig. 5.8(a). Experimental curves are scaled to match the theoretical peak height in the low V_{PN} range. Theoretical peaks were computed using SRH theory for recombination at interface traps located in the LDER, with the following parameters: mid-gap trap level E_{TI}=0; X_{OX}=4.5nm; $N_{XX}=5\times10^{19}cm^{-3}$ in the gate, and $N_{AA}=2.0\times10^{18}cm^{-3}$ in the LDER. The LDER concentration is determined from the theory-experimental match of peak voltage in the low V_{PN} curves.

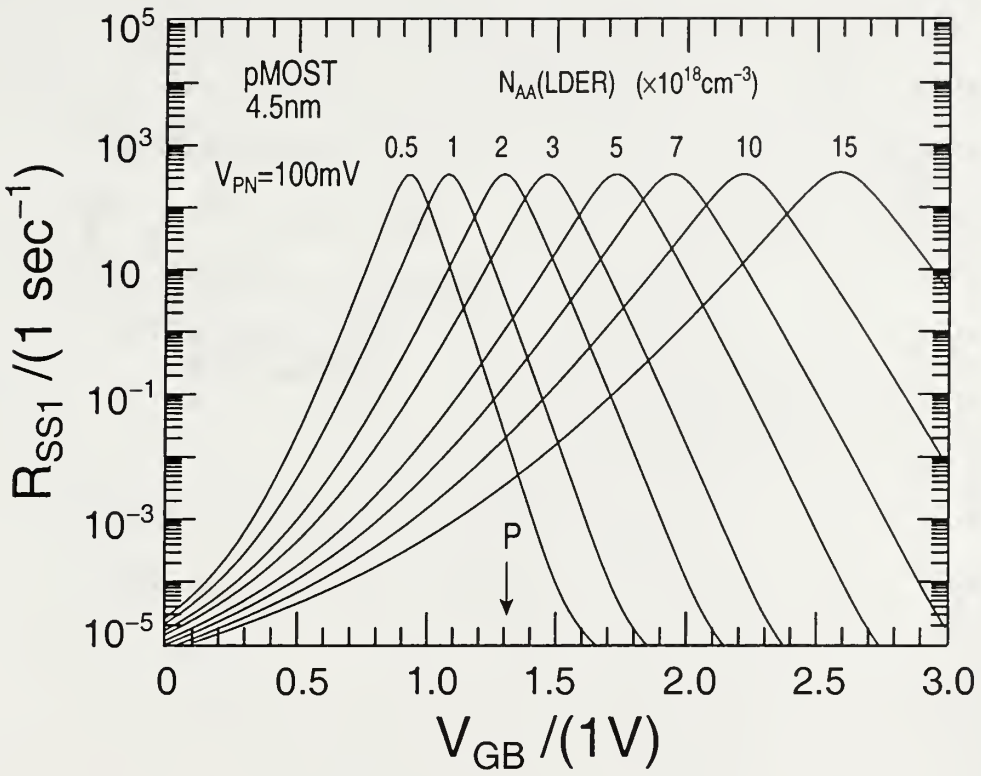

Fig. 5.12 Theoretical SRH recombination rate at interface traps in the gate overlapped drain region at a p+drain/n-base junction forward bias of 100mV, with the impurity doping concentrations in p+ drain/source LDER as a parameter. Locations of the I_B peak in the 20μm/0.25μm pMOST is indicated by an arrow. $N_{AA}=5\times10^{19}cm^{-3}$ in p+ gate and X_{OX}=4.5nm.

The experimental I_B curves in Fig. 5.11 were scaled to match the R_{SS1} peak at V_{PN}=100mV. The scaling factor provides an estimate on interface trap density N_{IT} in the overlap region. The number of interface traps in the overlap LDER area of the pMOST is given by

$$(W \times L_{overlap}) N_{IT} = \Delta I_{B\text{-}pk} / q R_{SS1\text{-}pk} \qquad (5.17)$$

$$= 13fA / (1.6 \times 10^{-19} C \times 325 s^{-1}) = 2.5 \times 10^{2}$$

where $L_{overlap} \approx 0.1\mu m$ is the length of overlap region, W=20μm is the channel width, 13fA is the peak current at V_{PN} = 100mV. The result of (5.17) normalized by the overlap area gives $N_{IT}=1.3\times10^{10}cm^{-2}$ in the LDER, which is higher than the typical 2-$8\times10^{9}cm^{-2}$ measured in the MCR on prestress MOSTs [5,33].

Though theory-experimental match in Fig. 5.11 is almost perfect for the peak voltage and peak current in the low V_{PN} range from 100mV to 200mV, deviations exist at higher V_{PN}'s. I_B peaks at 300mV and 350mV appear much higher than their theoretical predictions. The agreement will be significantly improved if the V_{GB}-independent baseline is subtracted from each of these two experimental curves. The baseline has the $\exp(qV_{PN}/kT)$ dependence while the LDER peak has the $\exp(qV_{PN}/2kT)$ from the ideal SRH theory, thus the baseline becomes less significant at lower V_{PN}'s to give better theory-experimental match. The experimental peak voltage shift 100mV more in the negative V_{GB} direction than the theoretical prediction as V_{PN} is increased from 100mV to 350mV. We ascribe this discrepancy to two-dimensional effects such as the fringe field in the LDER not considered by the theory.

Similar experimental characteristics on thicker oxide nMOST were reported by Acovic and co-workers [117] who measured forward-biased drain current as a function of gate voltage at 77K. The overall increase of drain current with negative gate voltage was attributed to interface trap assisted tunneling. Their sample has a thick oxide of 38nm, and there is a primary peak in the drain current that shifts from V_{GB}=−14V at V_{PN}=0.575V to V_{GB}=−12V at V_{PN}=0.975. The direction of the shift is consistent with interface traps in the gate overlapped LDER. Using the SRH recombination model just described, we found that a drain doping concentration of $2.2\times10^{19}cm^{-3}$ would perfectly match the peak location, which is not inconsistent with the high source/drain implantation dose of $10^{16}cm^{-2}$ (120keV, As) used in their sample.

5.5 Interface-Trap Assisted Tunneling

The rising $I_{B\text{-}acc}$ can not be explained by direct surface p/n junction tunneling due to small potential barrier at the p/n junction surface, while the $I_{B\text{-}acc}$ peaks provide evidence of interface traps in the gate overlapped lowly-doped (drain) extension region (LDER). In this section, we will consider interface trap assisted tunneling (TAT) in the LDER. To enable efficient computation using one-dimensional analytical formulas, we restrict the tunneling in the x-direction perpendicular to the Si/SiO_2 interface. In general, tunneling would occur along the direction of electric field at the surface, which is not limited to the x-direction near the junction space charge region where the field distribution is two-dimensional.

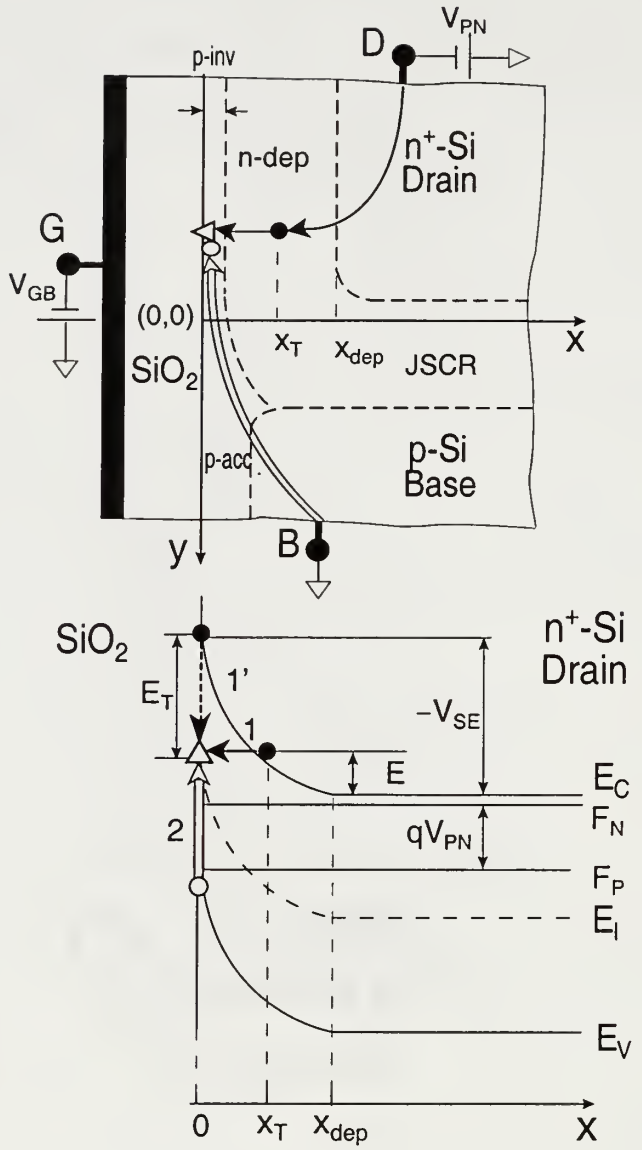

Fig. 5.13 Cross-sectional view and energy band diagram illustrating interface-trap assisted tunneling along the x-direction at the gate-overlapped n+ drain region. At large negative gate voltages, there is an electron depletion layer (thickness x_{dep}) and even a hole inversion layer at the surface of n+ drain, while there is a hole accumulation layer at the surface of p-base. The energy band diagram shows three processes: (1) Conduction band electron at $x=X_T$ tunnel into interface trap (with energy E_T below E_C) at x=0, (1') Thermal capture of a conduction band electron by the trap at x=0, and (2) Thermal capture of a valence band hole by the trap at x=0. F_P, F_N are electron and hole quasi Fermi levels, V_{PN} is the forward junction bias and V_{SE} denotes the surface band bending in n+Drain.

The trap-assisted tunneling pathway is illustrated in Fig. 5.13 with a cross-sectional view of the gate-overlapped n+D/p-Base junction and an energy band diagram along the x-direction. Under a large negative gate bias V_{GB} and a forward junction bias V_{PN}, n+D surface is electron depleted (n-dep) and has a hole inversion layer (p-inv). The hole inversion layer is connected to the hole accumulation layer (p-acc) at the surface of p-Base. The triangle symbolizes an interface trap in the n+D (LDER) with a discrete energy level E_T below the conduction band edge E_C. The interface trap can either thermally capture a conduction band (CB) electron at the surface x=0 (process 1') or elastically capture a CB electron at $x=X_T$ through quantum mechanical tunneling (process 1). Energy conservation requires that band bending at X_T equals $V_{SE} - E_T/q$, where V_{SE} is the total surface band bending in the x-direction. The relative probabilities of these two processes depend on the gate voltage. Process 2 is the thermal capture of a valence band hole at X=0, which is supplied by the base-terminal through the p-Base hole accumulation layer and the n+Drain hole inversion layer. The combination of processes 1' and 2 is the usual SRH recombination process. At more negative V_{GB}, electron concentration at x=0 drops exponentially which decreases SRH recombination rate, while the probability of tunneling (process 1) increases exponentially due to a thinner tunneling barrier. The combination of processes 1 and 2 is the interface-trap assisted tunneling process, which should dominate at large negative V_{GB}.

We will next develop a theory for surface p/n junction current that takes into account of all three processes (1,1'and 2). We start with the steady-state condition:

$$\partial n_{IT}/\partial t = 0 = \partial n_{IT}/\partial t(SRH) + \partial n_{IT}/\partial t(tunnel)$$

$$= (c_{ns}n_s p_{IT} - e_{ns}n_{IT}) - (c_{ps}p_s n_{IT} - e_{ns}n_{IT}) + J_T/q \qquad (5.18)$$

n_{IT} and p_{IT} are the electron-occupied and electron-depleted interface trap areal densities respectively, and $n_{IT} + p_{IT} = N_{ITT}$ is the total density of interface traps. n_S and p_S are the surface electron and hole densities, and c_{ns}, c_{ps}, e_{ns}, e_{ps} are electron and hole capture and emission rate coefficients respectively at the surface. The last term on the R.H.S. of (5.18), comes from electron tunneling from the conduction band at $x=X_T$ to interface traps at $x=0$ and J_T denotes the tunneling current density.

We then derive a formula for J_T based on (5.2) and the DOS of the final state (interface trap) in the tunneling process, which is $D_2(E)= p_{IT}\,\delta(E+qV_{SE}+E_T)$, where E is the electron energy of the initial conduction band state at $X=X_T$. The delta function reflects the conservation of electron energy. Since the tunneling probability is proportional to the density of final states, we apply a dimensionless multiplication factor to the integrand of (5.2) to reflect the specific DOS for tunneling into interface states. The result is:

$$J_T = \frac{4\pi\eta q\sqrt{m_y m_z}}{h^3}\int_0^{-qV_{SE}} T(E)\cdot[f(E)-f(E+qV_{PN})]E\cdot\frac{P_{IT}\,\delta(E+qV_{SE}+E_T)}{m_0/(\pi\hbar^2)}dE \qquad (5.19)$$

$$\equiv q\,p_{IT}\,T_0 \qquad (5.20)$$

$$T_0 = \eta\sqrt{m_y m_z}/(m_0 h)\cdot\{T(E)[f(E)-f(E+qV_{PN})]E\}\,|\,E=-qV_{SE}-E_T \qquad (5.21)$$

T_0 in (5.21) is the tunneling rate per second, which is evaluated at $E=-qV_{SE}-E_T$. It is mainly determined by two factors: (1) T(E), the barrier transmission probability, which increases with accumulation $|V_{GB}|$ due to barrier thinning and (2) f(E), the

electron occupation factor at $x=X_T$ which decreases with $|V_{GB}|$. To evaluate T(E), we use the average-field non-parabolic formula of (5.9) and replace the barrier height E_g by E_T. The average field in the tunneling region is: $F_{avg} = [F(x=0)+F(x=X_T)]/2$. The electric fields can be evaluated using the standard one-dimensional MOS equations.

The steady-state electron-occupied interface trap density can be derived from (5.18) and (5.20) in terms of T_0, which gives:

$$\frac{N_{IT}}{N_{ITT}} = \frac{c_{ns}N_s + e_{ps} + T_0}{c_{ns}N_s + c_{ps}P_s + e_{ns} + e_{ps} + T_0} \qquad (5.22)$$

Equation (5.22) reduces to the familiar SRH formula when the tunneling rate is much lower than the thermal capture rates. In the other limit when the tunneling rate dominates, the trap (electron) occupation factor approaches unity.

Combining (5.18) and (5.22), we derive the following formula that describes steady-state current follow due to all three processes (1, 1' and 2) at the surface of the LDER:

$$J_{LDER} = q\, N_{ITT}\, \frac{(c_{ns}N_s + T_0)\, c_{ps}P_s - e_{ns}e_{ps}}{c_{ns}N_s + c_{ps}P_s + e_{ns} + e_{ps} + T_0} \equiv qN_{ITT}R_1 \qquad (5.23)$$

R_1 is defined as a generalized recombination rate per interface trap, which reduces to the SRH rate R_{SS1} if the tunneling process is not considered. When the tunneling rate T_0 (process 1) is much higher than the thermal electron capture rate $c_{ns}N_s$ (process 1'), (5.23) is simplified to:

$$J_{LDER} = q\, N_{ITT}\, \frac{T_0 c_{ps}P_s - e_{ns}e_{ps}}{c_{ps}P_s + e_{ns} + e_{ps} + T_0} \qquad (5.24)$$

Equation (5.24) is the formula for trap-assisted tunneling (processes 1 and 2). For near mid-gap interface trap energy levels, the following equalities hold: $e_{ns} \approx c_{ns} n_i$ and $e_{ps} \approx c_{ps} n_i$. Therefore e_{ns} and e_{ps} are both negligible compare to $c_{ps}P_s$ and T_0. In the limit of strong accumulation such that $c_{ps}P_s >> T_0$, (5.24) can be further simplified to:

$$J_{LDER} = q\, N_{ITT} \frac{T_0 c_{ps} P_s}{c_{ps} P_s + T_0} \approx qN_{ITT}T_0 \qquad (5.25)$$

which describes the TAT-limited LDER current.

We will next examine the effects of gate bias, forward junction bias, n+D doping concentration, gate oxide thickness, interface trap energy level and temperature on the LDER current using the most general formula (5.23).

The gate voltage dependence of normalized LDER current ($R_1 = J_{LDER}/qN_{ITT}$) is shown in Fig. 5.14, along with that of tunneling rate T_0 [eqn (5.21)], the electron thermal capture rate $c_{ns}N_S$ and the hole thermal capture rate $c_{ps}P_S$ for a n+D/p-sub junction forward biased at 200mV. Parameters used in the computation include X_{OX}=4.5nm, N_{DD}(Drain)=$10^{18}cm^{-3}$ and N_{DD}(gate)=$5\times10^{19}cm^{-3}$, $E_T = E_g/2$ and T=300K. The gate voltage range (−3.0V to 0V) can be divided into three regions: (1) Tunneling current limited region at $V_{GB} < -1.6V$, where $c_{ps}P_S > T_0 > c_{ns}N_S$ and $R_1 = T_0$; (2) SRH recombination current limited region at $V_{GB} > -1.45V$, where $c_{ps}P_S$, $c_{ns}N_s > T_0 \approx 0$, and $R_1 = R_{SS1}$; (3) A narrow transition region around $V_{GB} = -1.5V$ in which T_0 and $c_{ns}N_s$ are compatible. The peak at $V_{GB} = -1.1V$ is from SRH recombination at interface traps in the LDER, which coincide with the point where $c_{ps}P_S = c_{ns}N_S$. In the second region, T_0 drops to zero since the surface band bending is not large enough to provide a direct

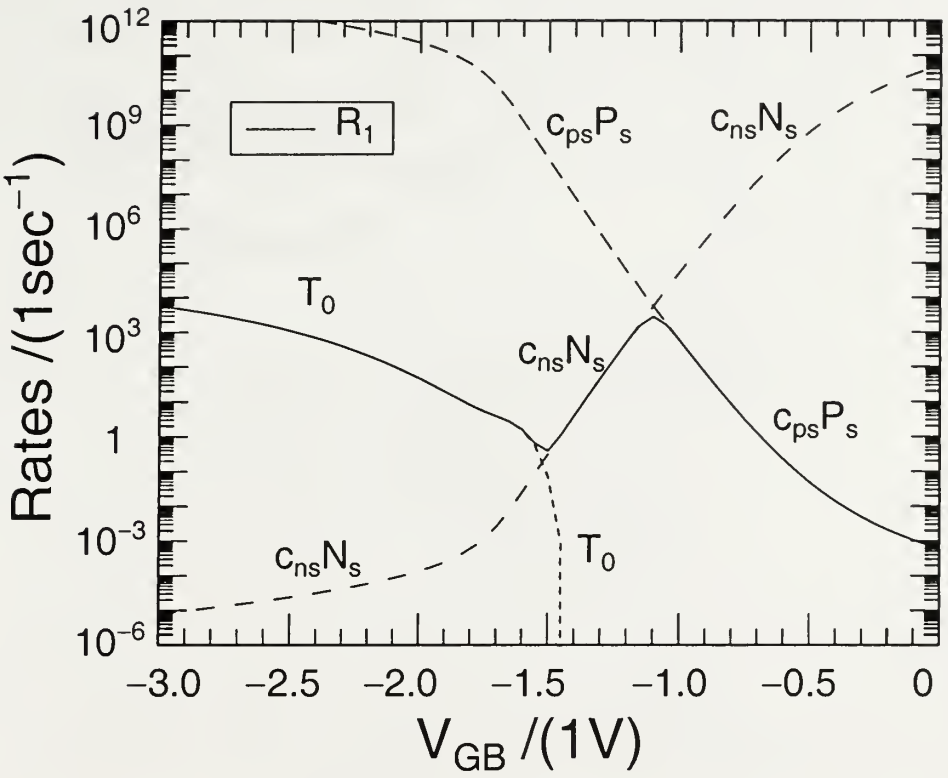

Fig. 5.14 Gate voltage dependence of the generalized recombination rate R_1 in the gate-overlapped drain region under a forward bias of 200mV on the n+drain/p-base junction. Three rates associated with interface traps are also shown which include electron and hole thermal capture rates $c_{ns}N_s$ and $c_{ps}P_s$, and tunneling rate T_0. Parameters used are: $N_{DD}=10^{18}cm^{-3}$ in n+ drain, $N_{DD}=5\times10^{19}cm^{-3}$ in n+ gate, $X_{OX}=4.5nm$, $E_{TI}=0$ and T=298K.

band-to-trap tunneling pathway. In the first region where the LDER current is limited by tunneling rate, the LDER current increases with stronger accumulation gate voltage, which is the characteristic of $I_{B\text{-}acc}$ data from DCIV measurement.

Figure 5.15 consists of four parts, each part shows a family of R_1-V_{GB} curves with one of the four variables as a parameter: (a) the forward junction bias V_{PN}; (b) doping concentration in n+Drain N_{DD}; (c) gate oxide thickness X_{OX} and (d) interface trap energy level E_{TI}. The solid line in each part is the same curve analyzed in Fig. 14 which serves as a reference. To enable lineshape comparison, the dashed curves in Figs. 5.15(b) and 5.15(c) are shifted along the V_{GB} axis to yield the same SRH peak at −1.1V. The V_{PN} dependence in Fig. 5.15(a) is in qualitative agreement with the experimental DCIV data in the strong accumulation range (Fig. 5.8): the TAT current increases with V_{PN} at a fixed V_{GB}. However, the LDER current increases faster with V_{PN} than experimental $I_{B\text{-}acc}$ data. In terms of ideality factor, the TAT theory gives n=1.27 V_{GB}=−3V, while the DCIV data gives n≈2 (Fig. 5.9). Figure 5.15(b) shows that the tunneling-limited LDER current at V_{GB}<−1.5V increases at higher n+D doping concentrations. The peaked structure near the onset of tunneling region is due to the interplay between (1) the number of electrons available for tunneling at x=X_T, which decreases with $|V_{GB}|$ and (2) the barrier transmission probability T(E), which increases with $|V_{GB}|$. The peak widening at higher N_{DD} concentrations is due to larger $|V_{GB}|$ required to invert the n+D surface. Fig. 5.15(c) shows that thin gate oxide leads to high LDER current in the tunneling-limited range, which is consistent with DCIV experiments that the rising I_B in accumulation range is more significant in thin oxide

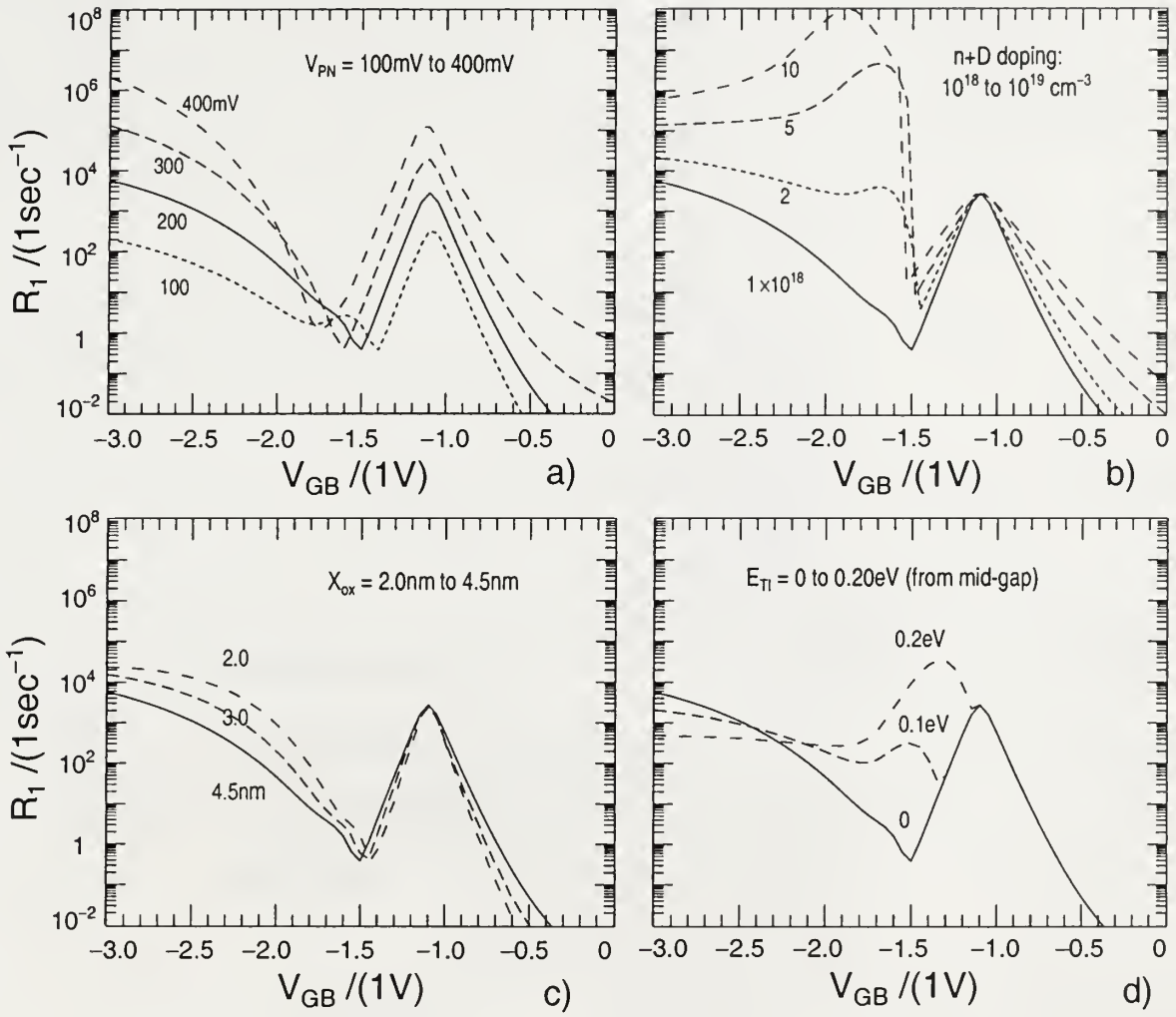

Fig. 5.15 Gate voltage dependence of the generalized recombination rate R_1 at **(a)** four junction forward biases, **(b)** four n+drain impurity doping concentrations, **(c)** four gate oxide thicknesses, and **(d)** three interface trap energy levels. The solid line is the reference curve with V_{PN}=200mV, N_{DD} =$10^{18}cm^{-3}$ X_{ox}=4.5nm and E_{TI}=0. In each part, only one parameter is varied from the reference value. In part a), the ideality factor n=1.27 at V_{GB}=−3.0V and n=2.0 for the peaks. In part b) and c), the dashed lines are shifted to match the peak voltage of the solid line at −1.10V.

MOSTs. The effect of interface trap energy level is illustrated in Fig. 5.15(d). A mid-gap trap has E_{TI}=0. For trap levels closer to the silicon conduction band edge, the threshold for tunneling is lowered due to a lower tunneling barrier ($E_T = E_g/2 - E_{TI}$), a peaked structure appears after the onset of tunneling, and the tunneling rate rises slower with $|V_{GB}|$ in the strong accumulation range.

Theoretical LDER currents at three different temperatures are shown in Fig. 5.16. The same parameters as those in 5.14 were used in the computation. Figure 5.16(a) shows unscaled R_1-V_{GB} curves which exhibits a significant temperature dependence even in the tunneling-limited range, in contrast to the notion that tunneling current has only a weak temperature dependence. This is because the tunneling probability is exponentially dependent on the tunneling distance X_T, which is not fixed in TAT. X_T has a weak temperature dependence through the MOS capacitor equation. In order to compare the relative strength of temperature dependence in SRH range and tunneling range, we scaled down the 80°C curve and scale up the −30°C curve in Fig. 5.16(b) to yield the same SRH peak as the room temperature curve. In the tunneling range on the negative V_{GB} side of the SRH peak, the −30°C curve is above the 80°C curve, which means the tunneling rate has a 'weaker' temperature dependence than that of the SRH recombination peak. On the other side of the SRH peak, the −30°C curve is below the 80°C curve, since in this range $R_1 = c_{ps}P_S \propto n_i^2 \exp(qV_{PN}/kT)$ has a stronger temperature dependence than $R_1 \propto n_i \exp(q_{PN}/2kT)$ at the peak.

In conclusion, we have developed a simple trap-assisted tunneling (TAT) model to account for majority carrier tunneling into interface traps located in gate-overlapped

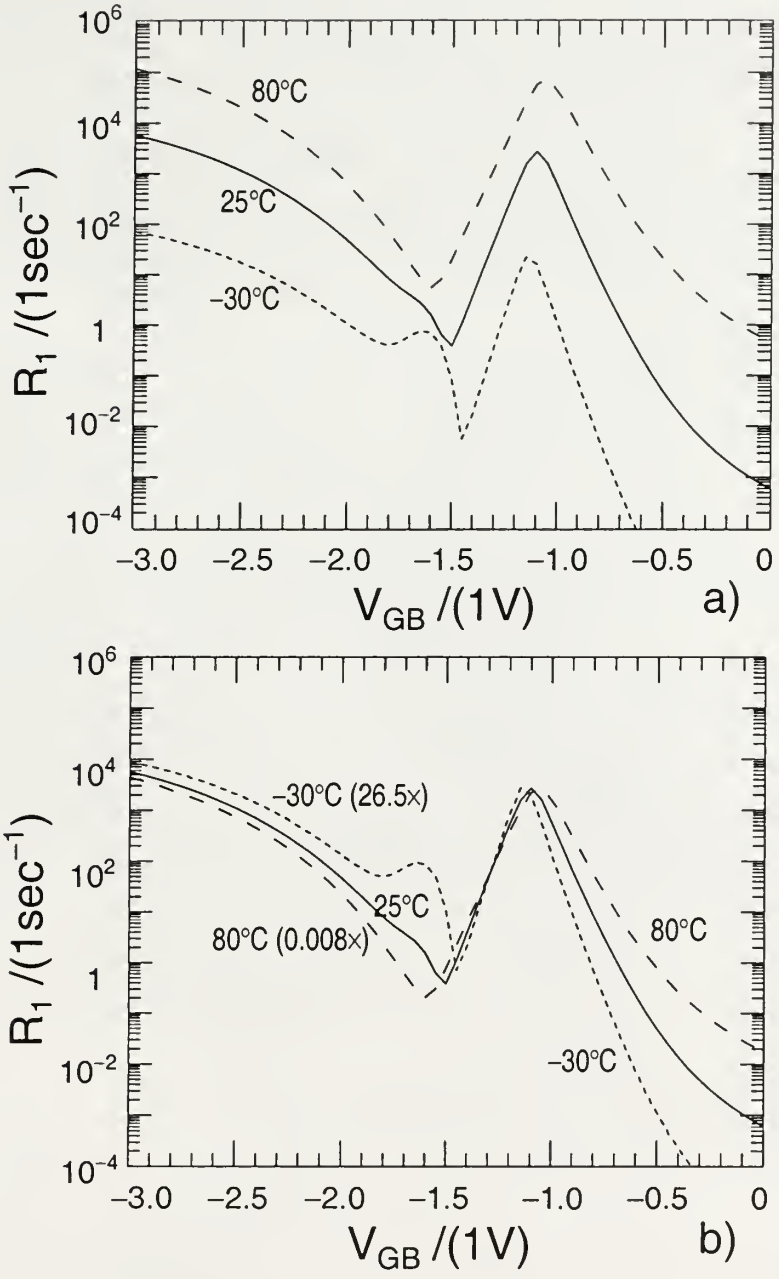

Fig. 5.16 Gate voltage dependence of the generalized recombination rate in the LDER at three temperatures: −30°C, 25°C and 80°C. **(a)** The three unscaled R_1 curves. **(b)** Two scaled scaled R_1 curves along with the unscaled room temperature curve. The scaling in part (b) matches the peak levels at −1.10V. Parameters other than temperature are the same as those in Fig. 5.14.

drain or source regions (LDER) of a MOS transistor. The model predicts a lower tunneling threshold $|V_{GB}|$ than that of the drain junction perimeter direct tunneling (DT) theory. The TAT rate increases with both MOST accumulation gate voltage and drain junction forward bias, which are the essential features of the DCIV base-terminal current (I_B) in accumulation V_{GB} range. However, the V_{PN} dependence of TAT current in the LDER is not in quantitative agreement with DCIV data, and some DCIV data such as those in Fig. 3.10 show no apparent I_B peaks in the accumulation V_{GB} range. Two sources may contribute to this discrepancy: (1) the simple tunneling theory does not consider the detailed interaction potential between the interface trap and the tunneling electron, and (2) the tunneling direction is not strictly perpendicular to the interface and therefore the purely one-dimensional theory fails to give the quantitative correlation.

We have also considered in detail another possible mechanism not related to interface traps which is the injection current in the LDER region. In this scenario, the V_{GB} dependence of the injection current comes from the gate voltage modulation of the series resistance in the surface (weak) inversion layer of the LDER. The theory of carrier injection over the inverted channel area of a gated p/n junction was first considered by Sah [3]. However, the following characteristics of the LDER injection current fails to correlate with experimental DCIV data: (1) The injection current would quickly saturate at strong accumulation gate voltages when the LDER surface is strongly inverted while experimental $I_{B\text{-}acc}$ data has no indication of saturation; (2) The LDER injection current is independent of interface trap density while the experimental $I_{B\text{-}acc}$ increases after channel-hot-carrier stress indicating a N_{IT} related mechanism.

5.6 Summary

Three original theories of surface controlled interband tunneling near the drain/substrate junction perimeter in MOS transistors under forward junction bias have been developed: (1) electron tunneling from n+drain into classical three-dimensional hole states in p-substrate surface, (2) electron tunneling from n+drain into quantized two-dimensional states of surface hole accumulation layers, and (3) interface trap assisted tunneling (TAT) in the gate-overlapped drain extension region. The first two theories predict a high threshold gate voltage which increases as V_{PN} increases, not consistent with DCIV data. The third theory has lower threshold gate voltage and are in qualitative agreement with the experimental DCIV data in the MOST's accumulation gate voltage range which show I_B peaks on a rising baseline. At a given V_{GB}, however, the experimental I_B data ($n \approx 2$) rises with V_{PN} slower than that from the TAT theory ($n=1.27$ at $V_{GB}=-3V$). The quantitative correlation may be improved by taking into account of lateral voltage drop in the (weakly) inverted LDER surface, which was shown by Sah [3] to give a factor of two increase in the ideality factor in the limit of a long surface inversion channel. The DCIV peaks in the accumulation gate voltage were correlated to SRH recombination at the interface traps located over the LDER region.

CHAPTER 6
SUMMARIES AND CONCLUSIONS

The success of today's semiconductor industry can be partially attributed to the passivation of intrinsic defects at the interface between the silicon channel and the thermal gate oxide to give extremely-low electrically-active interface trap density on as-manufactured MOS transistors. The direct-current current voltage (DCIV) method has the sensitivity to detect these low-density interface traps by measuring the gate bias (V_{GB}) and the forward junction bias (V_{PN}) modulated electron-hole recombination currents from the base-terminal (I_B). This unique sensitivity of DCIV makes it a powerful tool for monitoring the transistor reliability and for diagnosis of transistor design.

In previous chapters, fundamental theories which relate DCIV characteristics to device and material parameters were presented. It is shown that the DCIV peak current, peak gate voltage and lineshape can each be used to extract a group of parameters: interface trap density and energy levels from the V_{PN} dependence of DCIV peak currents; spatial profile of interface traps from DCIV lineshape and channel length dependence of DCIV peak current; and impurity doping concentration and gate oxide thickness from the V_{PN} dependence of DCIV peak gate voltages. These applications were demonstrated on both the unstressed MOS transistors as well as hot-carrier-stressed MOS transistors. Diffusion limits the minority carriers injected from the

bottom-emitter to reach at the Si/SiO_2 interface at low bottom emitter biases which produces a wider and lower peak current than the theoretical prediction. The interface trap energy levels are discrete rather than continuous, consistent with those of silicon dangling bonds. Spatial density of interface trap over the channel area of unstressed MOS transistors is shown to rise sharply towards the source and drain junctions with a characteristic length of 80nm. Increased interface trap density was observed in the drain junction space charge region after channel-hot-carrier stress. Transistor design parameters (doping concentration at the interface and gate oxide thickness) extracted from DCIV characteristic are in good agreement with designed values.

Transistor characterization becomes more difficult on ultrathin oxide MOS transistors in current and future technology generations due to the interference of gate oxide tunneling current which appears as a rising baseline in DCIV base-terminal current. Prompted by the symmetry of the gate tunneling current-voltage (I_G-V_{GB}) characteristics in p-channel and n-channel MOS transistors, we developed a first-principle gate oxide tunneling theory that shows hole tunneling is dominant in p+gate p-channel MOS transistors and that the small gate/drain overlap area is not negligible compare to the channel area which gives the knee point observed in the I_G-V_{GB} curves of both nMOST and pMOST. The theory improves the traditional WKB barrier transmission probability by taking into account of quantum mechanical wavefunction matching at the abrupt potential boundaries. This improvemnet is critical in accounting for the experimental data at low electric fields. We demonstrated the separation of two geometric tunneling pathways in gate currents and the extraction of impurity doping

concentrations in the channel and drain/source areas from the theory-experimental correlation.

In the gate accumulation voltage range of thin oxide MOS transistors, the current flowing through the drain/base p/n junction increases with gate voltage, which is another source of rising baseline in DCIV curves. We developed two fundamental theories of interband tunneling to compare with the experimental data: (1) direct tunneling at the drain/base junction perimeter and (2) interface trap assisted tunneling in the gate-overlapped drain region. Energy quantization in the surface accumulation layer of the p-base is shown to give breakpoints (not peaks) in the direct tunneling current at the gate voltages when one of the one-dimensional discrete energy levels crosses the conduction band edge on the n+drain side. Direct tunneling theory predicts a high threshold gate voltage which increases as V_{PN} increases, not consistent with DCIV data. Interface trap assisted tunneling has lower threshold gate voltage and can qualitatively explain the V_{PN} dependence of the experimental I_B baseline. In particular, the existence of interface traps in the overlap region is confirmed by experimental I_B peaks at gate voltages consistent with the designed drain impurity doping concentrations. Better theory-experimental agreement entails numerical two-dimensional field analysis near the drain/base junction perimeter.

Tunneling current will eventually limit the further thinning of gate oxide below 1.5nm. Alternative high dielectric constant gate materials are being actively sought. Both the DCIV method and tunneling current monitor investigated in this thesis research

will continue to be useful for evaluation of interface reliability as well as for diagnosis of transistor design in the next generations of semiconductor technologies.

APPENDIX
DERIVATION OF GATE OXIDE TUNNELING FORMULAE

The appendix presents a derivation of the gate oxide tunneling formulae used in chapter 4, following Bardeen's transition probability approach [59] and Harrison's independent-particle tunneling model [60]. An alternative approach using barrier transmission coefficient (TC) would give the same result if wavefunction matching at the abrupt potential boundaries are taken care of when deriving the transmission coefficient. However, the traditional TC method uses a simple exponential transmission coefficient with unity pre-exponential factor which neglected the wavefunction matching [53].

Wavefunctions of the Initial and Final States

Consider a trapezoidal SiO_2 barrier and two electron states: a right-hand state Ψ_{S1} and a left-hand state Ψ_{S2}. Ψ_{S1} and Ψ_{S2} are the quantum mechanical wavefunctions of the two states. Because of the one-dimensional potential barrier [V(x)], the transverse part of the two wavefunctions are trivial and can be both represented by $\exp(ik_y y+ik_z z)$. The same wave numbers (k_y,k_z) in the initial and final states are a result of transverse momentum conservation. In the following analysis, we will focus on the normal part of the wavefunctions Ψ_{S1X} and Ψ_{S2X}. For notation compactness, we will drop the subscript X, so Ψ_{Si}, k_i and E_i will be used to represent the x-component of the wavefunction, wave vector and kinetic energy respectively in the initial state (i=1) and

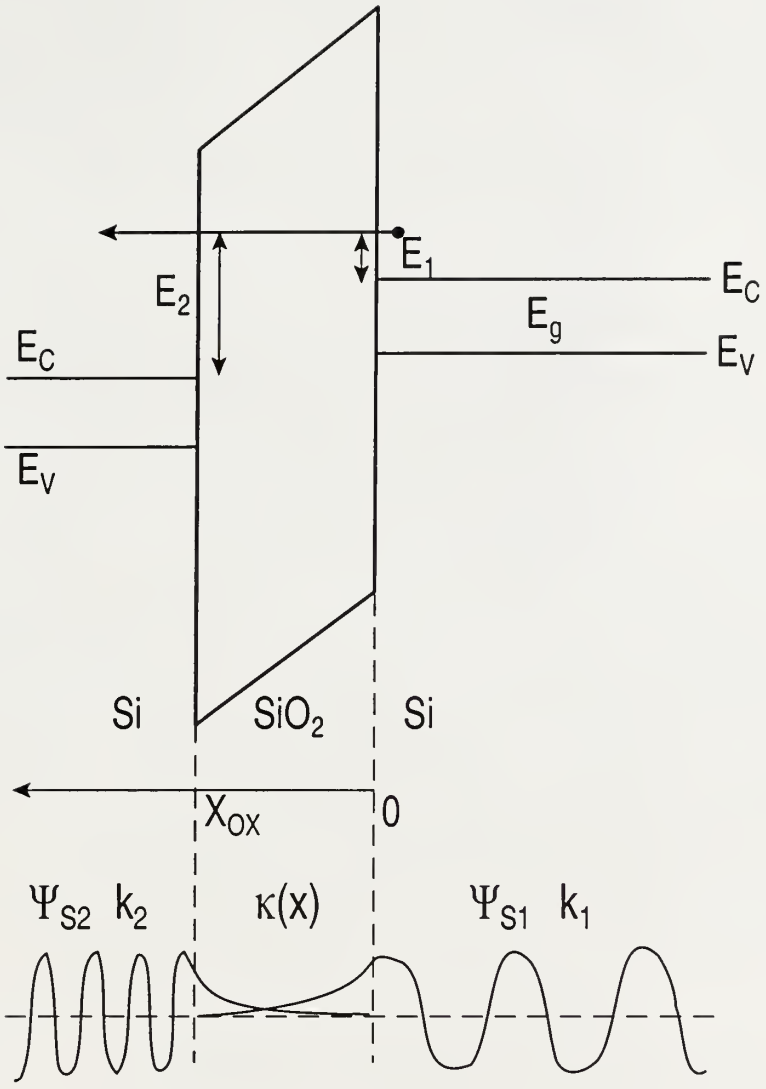

Fig. A.1 Energy band diagram of a trapezoidal SiO_2 barrier and wavefunctions of the initial (right-hand) state and the final (left-hand) state of an elastic tunneling transition process. k_1 and k_2 ($k_2>k_1$) are the real wavenumbers outside of the barrier region and $\kappa(x)$ is the imaginary wavenumber inside the barrier region ($0<x<X_{OX}$).

the final state (i=2). Ψ_{S1} and Ψ_{S2} are sinusoidal at x>0 and at $x<X_{OX}$ respectively, while they decay exponentially in the opposite directions inside the potential barrier, as shown in Fig. A.1.

From Bardeen's transition probability point of view, tunneling of an electron from the right-side of the barrier to its left-side is equivalent to the transition of an electron in an initial state Ψ_{S1} to a final state Ψ_{S2}. With the free-electron approximation outside the barrier region and the WKB approximation inside the barrier region, the wavefunctions would be of the form:

$$\Psi_{S1} = A_1 \cos(k_1 x + \Delta_1) \qquad (x<0) \qquad (A.1a)$$

$$= \frac{B_1}{\sqrt{\kappa(x)}} \exp\left[-\int_0^x \kappa(x)\,dx\right] \qquad (0<x<X_{OX}) \qquad (A.1b)$$

$$\Psi_{S2} = A_2 \cos[k_2(x-X_{OX}) + \Delta_2] \qquad (x>X_{OX}) \qquad (A.2a)$$

$$= \frac{B_2}{\sqrt{\kappa(x)}} \exp\left[-\int_x^{X_{OX}} \kappa(x)\,dx\right] \qquad (0<x<X_{OX}) \qquad (A.2b)$$

k_1 and k_2 are the wavenumbers outside the barrier region in the initial and final states respectively, and $\kappa(x)$ is the imaginary wavenumber inside the barrier which is given by eqn. (4.6). The various constants A_i, Δ_i and B_i (i=1,2) can be determined from (1) normalization of wavefunctions and (2) wavefunction matching at the two abrupt potential boundaries. In the following, we will show the procedures to determine B_1 and B_2, which are the two constants pertinent to the evaluation of the transition matrix element.

Continuation of both the probability function and electron flux requires that both Ψ_S and $(1/m)\partial\Psi_S/\partial x$ be continuous across the abrupt boundaries. Applying these boundary conditions at x=0 for Ψ_{S1} gives:

$$\Psi_{S1}(x=0-) = A_1\cos\Delta_1 = \Psi_{S1}(x=0+) = B_1/\sqrt{\kappa(0)} \tag{A.3}$$

and

$$(1/m_{s1})\partial\Psi_{S1}/\partial x(x=0-) = -(k_1/m_{s1}\sin\Delta_1)A_1 =$$

$$(1/m_{ox})\partial\Psi_{S1}/\partial x(x=0+) = -(\sqrt{\kappa(0)}/m_{ox})B_1 \tag{A.4}$$

Elimination of Δ_1 from (A.3) and (A.4) results in the following expression for the magnitude of B_1:

$$|B_1| = |A_1|\times\left[\frac{1}{\kappa(0)} + \frac{m_{s1}^2\,\kappa(0)}{m_{ox}^2\,k_1^2}\right]^{-1/2} \tag{A.5}$$

Repeating the above procedures for Ψ_{S2} at $x=X_{OX}$, we derive a similar expression for the magnitude of B_2:

$$|B_2| = |A_2|\times\left[\frac{1}{\kappa(X_{OX})} + \frac{m_{s2}^2\,\kappa(X_{OX})}{m_{ox}^2\,k_2^2}\right]^{-1/2} \tag{A.6}$$

A_1 and A_2 can be determined from the normalization of wavefunctions in the initial and final states respectively, while assuming a small and negligible amount of wavefunction penetration inside the barrier region:

$$1 = \int_{-L_{S1}}^{X_{OX}} |\Psi_{S1}|^2 dx \approx \int_{-L_{S1}}^{0} |\Psi_{S1}|^2 dx$$

$$= \int_{-L_{S1}}^{0} |A_1|^2\cos^2(k_1x+\Delta_1)\,dx = |A_1|^2(L_{S1}/2) \tag{A.7}$$

which gives

$$|A_1| = (2/L_{S1})^{1/2} \tag{A.8a}$$

and a similar normalization of Ψ_{S2} gives

$$|A_2| = (2/L_{S2})^{1/2} \tag{A.8b}$$

where L_{S1} and L_{S2} are the extensions of silicon structure to the right side ($x<0$) and to left side ($x>X_{OX}$) of the SiO_2 barrier respectively. Thus the wavefunctions of the initial and final states are determined except for the phase shift Δ_i which does not affect the tunneling transition probability.

Transition Matrix Element

Bardeen [59] showed that the transition matrix element in the tunneling probability expression of (4.1) is proportional to that of the current density operator in the tunneling direction (J_x), which is to be evaluated between the initial state Ψ_{S1} and the final state Ψ_{S2} in the barrier region $0<x<X_{OX}$. Using the results of (A.1b), (A.2b), (A.5) and (A.6), the matrix element is evaluated as follows:

$$M_{12} \equiv \langle \Psi_{S1} | M | \Psi_{S2} \rangle$$

$$= -i\hbar \langle \Psi_{S1} | J_x | \Psi_{S2} \rangle \tag{A.9a}$$

$$= (-i\hbar)(i\hbar/2m_{ox}) \int_0^{X_{OX}} [\Psi_{S1}(\partial\Psi_{S2}^*/\partial x) - \Psi_{S2}^*(\partial\Psi_{S1}/\partial x)]\,dx \tag{A.9b}$$

$$= (\hbar^2/2m_{ox})\ (2B_1B_2^*)\ \exp\left[-\int_0^{X_{OX}} \kappa(x)\,dx\right] \tag{A.9c}$$

$$= (\hbar^2/2m_{ox})\ (2A_1A_2^*)\ \exp\left[-\int_0^{X_{OX}} \kappa(x)\,dx\right] \times$$

$$\left\{\left[\frac{1}{\kappa(0)} + \frac{m_{s1}^2\ \kappa(0)}{m_{ox}^2\ k_1^2}\right]\left[\frac{1}{\kappa(X_{OX})} + \frac{m_{s2}^2\ \kappa(X_{OX})}{m_{ox}^2\ k_2^2}\right]\right\}^{-1/2} \tag{A.9d}$$

where Ψ_{Si}^*, A_i^* and B_i^* (i=1,2) are the complex conjugate of Ψ_{Si}, A_i and B_i respectively. In deriving (A.9c), we assumed that the x-dependence of Ψ_{S1} and Ψ_{S2} inside the barrier region is dominated by the exponential term.

The density-of-state (DOS) factor per unit energy interval on the two sides of the SiO_2 barrier can be derived from the DOS factor in the k-space:

$$\rho_1 = (L_{S1}/\pi) \cdot (\partial E_1/\partial k_1)^{-1} = (L_{S1}/\pi) \cdot (m_{s1}/\hbar^2 k_1) \qquad (A.10a)$$

$$\rho_2 = (L_{S2}/\pi) \cdot (\partial E_2/\partial k_2)^{-1} = (L_{S2}/\pi) \cdot (m_{s2}/\hbar^2 k_2) \qquad (A.10b)$$

The length (L_{S1} and L_{S2}) dependence disappears in the following dimensionless combined factor of tunneling probability which can be obtained from (A.8a), (A.8b), (A.9d), (A.10a) and (A.10b) with some simple algebra:

$$\rho_1\rho_2|M_{12}|^2 = 1/(2\pi)^2 \times \exp\left[-2\int_0^{X_{OX}} \kappa(x)\,dx\right] \times \frac{16}{\left[\frac{m_{ox}k_1}{m_{s1}\kappa(0)} + \frac{m_{s1}\kappa(0)}{m_{ox}k_1}\right]\left[\frac{m_{ox}k_2}{m_{s2}\kappa(X_{OX})} + \frac{m_{s2}\kappa(X_{OX})}{m_{ox}k_2}\right]}$$

$$\equiv 1/(2\pi)^2 \times \exp(-2\upsilon) \times T_0(E_1, E_2) \qquad (A.11)$$

T_0 and υ defined in (A.11) are the same as those given by (4.4) and (4.5) respectively. T_0 depends on electron kinetic energies in the initial state E_1 and in the final state E_2 via the relationship $k_i=(2m_{si}E_i)^{1/2}$ (i=1,2). The requirement of energy conservation gives $E_2=E_1-qV_{OX}$, where V_{OX} is the voltage drop across the SiO_2 layer.

Tunneling Current Formulae

In this section, we will give the details of deriving the tunneling current formulae (4.3), (4.10) and (4.11). We start from (4.2) and convert the summation over transverse

momentum (k_y and k_z) to a double integration using a DOS factor of $1/(2\pi)^2$ in the two-dimensional k-space:

$$J = \frac{4\pi q}{\hbar} \sum_{k_t} \int_0^{E_m} |M_{12}|^2 \, \rho_1 \rho_2 (f_1 - f_2) \, dE_1 \qquad (A.12)$$

$$= \frac{4\pi q}{\hbar} \int_0^{E_m} dE_{1x} \left[\int_{-\infty}^{\infty} dk_{1y} \int_{-\infty}^{\infty} dk_{1z} \frac{1}{(2\pi)^2} (f_1 - f_2) \right] |M_{12}|^2 \, \rho_1 \rho_2 \qquad (A.13)$$

f_1 and f_2 are electron occupation factors in the initial and final states respectively. They are described by the Fermi-Dirac distribution function:

$$f_i = \{\exp[E_{ix} + E_{it} + E_C - E_F)/kT] + 1\}^{-1} \qquad (i=1,2) \qquad (A.14)$$

In (A.14), the electron kinetic energy above the conduction band minimum E_C is decomposed into a normal part E_{ix} and a transverse part $E_{it}=E_{iy}+E_{iz}$, E_F is the Fermi-level in the silicon layer.

The double integration of f_i over the transverse momentum can be evaluated analytically since f_i only depends on the magnitude of the transverse momentum for the simplest spherical E(**k**) relationship. For simplicity, we drop the subscript 'i' in the following evaluation:

$$\int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_y \, f = 2\pi \int_0^{\infty} k_t dk_t \frac{1}{\exp[(E_x + E_t + E_C - E_F)/kT] + 1}$$

$$= \frac{2\pi m_d kT}{\hbar^2} \int_0^{\infty} \frac{d(E_t/kT)}{\exp[(E_x + E_t + E_C - E_F)/kT] + 1}$$

$$= (2\pi m_d kT/\hbar^2) \cdot \log_e\{1 + \exp[(E_F - E_C - E_x)/kT]\} \qquad (A.15)$$

Equation (A.15) is also valid for ellipsoidal E(**k**) relationship such as those of silicon conduction band valleys, with the 2-dimensional DOS effective mass $m_d=(m_x m_y)^{1/2}$.

The Fermi-levels in the initial and final state differ by qV_{GB}, where V_{GB} is the applied gate-to-substrate voltage bias. Thus, the [ ] term in (A.13) gives the following results:

$$[\ ] = \frac{2\pi m_d kT}{h^2} \log_e\left[\frac{1+\exp[(E_F-E_C-E_x)/kT]}{1+\exp[(E_F-E_C-E_x-qV_{GB})/kT]}\right] \qquad (A.16)$$

The electron tunneling current formula (4.3) is derived by using the results of (A.11) and (A.16) in (A.13), which gives:

$$J_N = \frac{4\pi m_d kT}{h^3}\int_0^{\phi_{Bn}} \exp[-2\upsilon(E_x)]T_0(E_x)\log_e\left\{\frac{1+\exp[(E_F-E_C-E_x)/kT]}{1+\exp[(E_F-E_C-E_x-qV_{GB})/kT]}\right\}dE_x \qquad (A.17)$$

The integration in (A.17) is over the allowed kinetic energy (x-component) range of the initial state from zero to $E_m=\phi_{Bn}$, where ϕ_{Bn} is the Si/SiO_2 conduction band offset.

The hole tunneling formula (4.11) can be derived in the same way by considering the silicon valence band hole states on the two sides of the SiO_2 barrier as the initial and final states and taking into account of all allowed hole kinetic energies from zero to $E_m=\phi_{Bp}$, where ϕ_{Bp} is the Si/SiO_2 valence band offset.

For valence electron tunneling, the initial state is a silicon valence band electron state on one side of SiO_2 barrier and the final state is a silicon conduction band electron state on the other side of the barrier. Elastic tunneling path opens when the valence band overlaps the conduction band, which requires a minimum oxide voltage drop of qE_g. Under such oxide voltages, the valence band states are fully occupied while the conduction band states on the other side of SiO_2 are nearly empty, which gives $f_1-f_2 \approx 1-0 = 1$. The amount of conduction band and valence band overlap determines the

allowed kinetic energy range for elastic tunneling $0<E_1=E_{1t}+E_{1x}<E_m=qV_{OX}-E_g$. The allowed range for the transverse part of kinetic energy is thus $0<E_{1t}<E_m-E_{1x}$ which sets an upper limit for the transverse momentum $k_t<k_m=[2m_d(E_m-E_{1x})/\hbar^2]^{1/2}$. The result of (A.11) for the transition matrix element applies to valence electron tunneling as well, where $E_1(k_1)$ and $E_2(k_2)$ are referenced to the valence band edge and conduction band edge respectively. Starting from the general tunneling current formula (4.2), we derive the valence electron tunneling formula (4.10) as follows:

$$J_{VE} = \frac{4\pi q}{\hbar} \sum_{k_t} \int^{E_m} |M_{12}|^2 \rho_1\rho_2(f_1-f_2)\, dE_1$$

$$= \frac{4\pi q}{\hbar} \int_0^{E_m} dE_x \left[\int_0^{k_m} dk_y \int_0^{\sqrt{k_m^2-k_y^2}} dk_z \frac{1}{(2\pi)^2} (f_1-f_2) \right] |M_{12}|^2\rho_1\rho_2 \qquad \text{(A.18a)}$$

$$= \frac{4\pi q}{\hbar} \int_0^{E_m} dE_x \left[\int_0^{k_m} 2\pi k_t dk_t \frac{1}{(2\pi)^2} (1-0) \right] \frac{1}{(2\pi)^2} \exp(-2\upsilon) T_0(E_x) \qquad \text{(A.18b)}$$

$$= \frac{4\pi m_d}{h^3} \int_0^{qV_{OX}-E_g} (qV_{OX}-E_g-E_x) \exp[-2\upsilon(E_x)] T_0(E_x)\, dE_x \qquad \text{(A.18c)}$$

Triangular Barrier Tunneling

For a triangular barrier illustrated in Fig. A.2, the right-hand initial state is described by the same wavefunction Ψ_{S1} of the trapezoidal barrier. The procedures of matching the wavefunction at x=0 demonstrated in (A.3) to (A.5) apply to the present case of triangular barrier as well. However, for the left-hand final state Ψ_{S2}, there is no discontinuity at the potential boundary x=X'. We may write Ψ_{S2} using the WKB approximation at x>X' and x<X' respectively:

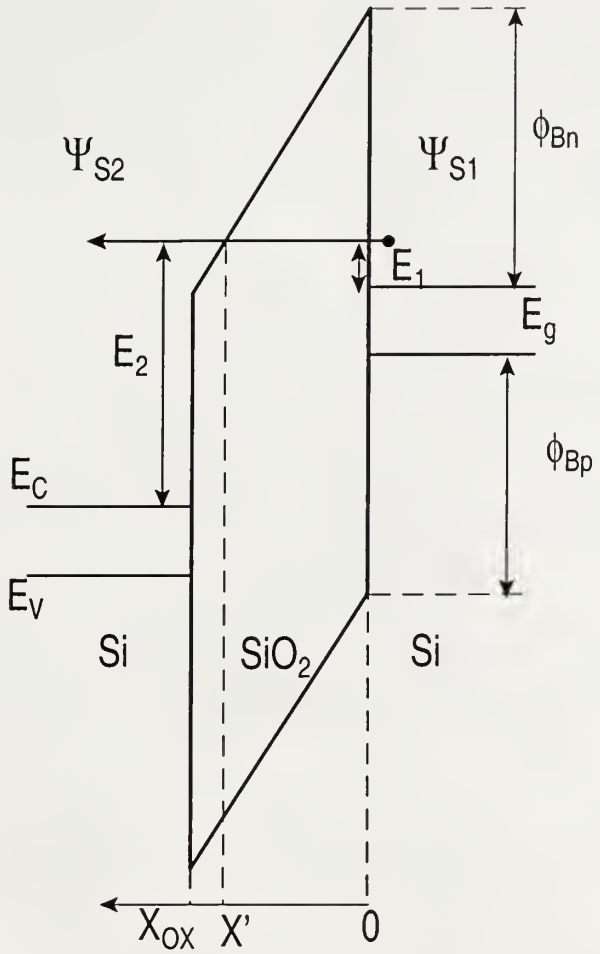

Fig. A.2 Energy band diagram of a triangular SiO_2 barrier. E_1 and E_2 are the kinetic energies referenced to the silicon conduction band minimum in the initial state Ψ_{S1} and final state Ψ_{S2} respectively. Two boundaries of the triangular barrier are at x=0 (abrupt) and x=X'(continuous).

$$\Psi_{S2} = \frac{A_2}{\sqrt{k_2(x)}} \cos\left[\int_{X'}^{x} k_2(x)dx + \pi/4\right] \qquad (x>X') \qquad (A.19a)$$

$$= \frac{A_2}{2\sqrt{\kappa(x)}} \exp\left[-\int_{x}^{X'} \kappa(x)dx\right] \qquad (0<x<X') \qquad (A.19b)$$

$k_2(x)$ is the local real wave number in the WKB approximation which becomes position-independent at $x>X_{OX}$. Note that (A.19a) and (A.19b) are related to each other by analytic continuation in a complex x-plane. Since $[\kappa(x)]^{1/2} \propto [V(x)-E]^{1/4} \propto (x-X')^{1/4}$, an excursion into the complex x-plane around x=X' picks up a phase $\pi/4$, which appears in (A.19a). $A_2 = (2k_2/L_{S2})^{1/2}$ is determined from normalization of Ψ_{S2}.

Both (A.19a) and (A.19b) become invalid near the classical turning point x=X' where $k_2(x)$ and $\kappa(x)$ reduce to zero. This is in contrast to Ψ_{S2} described by (A.2a) and (A.2b) for a trapezoidal barrier which are both valid near $x=X_{OX}$ due to the abrupt potential boundary. Using this approximate wavefunction of the final state inside the triangular barrier region, it would be straight forward to compute the transition matrix element and the tunneling current formulae using the same approaches demonstrated in previous sections. The result for the dimensionless tunneling transition probability is:

$$\rho_1\rho_2|M_{12}|^2 = \frac{1}{(2\pi)^2} \exp(-2\upsilon) \frac{4(m_{s2}/m_{ox})}{m_{ox}k_1/m_{s1}\kappa(0)+m_{s1}\kappa(0)/m_{ox}k_1}$$

$$\equiv \frac{1}{(2\pi)^2} \exp(-2\upsilon)\ T_0'(E_1) \qquad (A.20)$$

The only difference between the trapezoidal formula (A.11) and the triangular formula (A.20) is the pre-exponential factor. T_0 in (A.11) depends on kinetic energies in both the initial state (E_1) and the final state (E_2). This gives T_0 its V_{GB} dependence through

the relationship $E_2=E_1-qV_{OX}(V_{GB})$. In case of triangular barrier, T_0' in (A.20) only depends on E_1 and has no V_{GB} dependence.

In conclusion, we derived tunneling current formulae in the frame work of time dependent perturbation theory. The result shows a kinetic energy dependent pre-exponential factor T_0, while the simple barrier transmission coefficient in traditional formulae gives $T_0=T_0'=1$ [53]. For trapezoidal barrier, T_0 was derived rigorously and it gives a strong gate voltage dependence near the flatband condition which is indispensable to account for the experimental tunneling current-voltage characteristics presented in chapter 4 [Fig. 4.13(b)]. For triangular barrier, T_0' is less accurate due to the breakdown of WKB approximation near the potential boundary x=X' and T_0' is independent of gate voltage. Therefore in this high-field triangular barrier range, the traditional simpler tunneling formula with $T_0=T_0'=1$ can be used.

Summary of Detailed Tunneling Current Formulae

In the final section, we list the detailed tunneling current formulae for three tunneling carrier species (electron, valence electron and hole) and two shapes of tunneling barrier (trapezoid and triangle), which can be easily programmed to verify the theoretical curves given in chapter 4. These formulae give the explicit dependence on voltage drops in the silicon and oxide layers by considering (1) direction of tunneling in two gate voltage ranges $V_{GB}>V_{FB}$ and $V_{GB}<V_{FB}$ (2) the dependence of Fermi-level at the Si/SiO_2 interface on silicon band bendings and (3) parabolic $E(\kappa)$ relationship inside

the SiO_2 barrier. Silicon band structures are taken into account by effective mass approximations. The following formulae apply to both barrier shapes:

$$J_N = \frac{4\pi kT}{h^3}\int_0^{\phi_{Bn}} dE_x \exp[-2\upsilon(E_x)][4m_d^t T_0(E_x,m_s^t)+2m_d^l T_0(E_x,m_s^l)]$$

$$\times \log_e\left\{\frac{1+\exp[(E_{FG}-E_{CG}-E_x-qV_{SG}+qV_{GB})/kT]}{1+\exp[(E_{FG}-E_{CG}-E_x-qV_{SG})/kT}\right\} \quad (V_{GB}<V_{FB}) \qquad (A.21a)$$

or

$$\times \log_e\left\{\frac{1+\exp[(E_{FX}-E_{CX}-E_x+qV_{SX})/kT]}{1+\exp[(E_{FX}-E_{CX}-E_x+qV_{SX}-qV_{GB})/kT}\right\} \quad (V_{GB}>V_{FB}) \qquad (A.21b)$$

$$J_P = \frac{4\pi kT}{h^3}\int_0^{\phi_{Bp}} dE_x \exp[-2\upsilon(E_x)][m_d^{lh} T_0(E_x,m_s^{lh})+m_d^{hh} T_0(E_x,m_s^{hh})]$$

$$\times \log_e\left\{\frac{1+\exp[(E_{VX}-E_{FX}-E_x-qV_{SX}+qV_{GB})/kT]}{1+\exp[(E_{VX}-E_{FX}-E_x-qV_{SX})/kT]}\right\} \quad (V_{GB}<V_{FB}) \qquad (A.22a)$$

or

$$\times \log_e\left\{\frac{1+\exp[(E_{VG}-E_{FG}-E_x+qV_{SG})/kT]}{1+\exp[(E_{VG}-E_{FG}-E_x+qV_{SG}-qV_{GB})/kT]}\right\} \quad (V_{GB}>V_{FB}) \qquad (A.22b)$$

$$|J_{VE}| = \frac{4\pi m_d}{h^3}\int_0^{q|V_{OX}|-E_g} dE_x \exp[-2\upsilon(E_x)][m_d^{lh} T_0(E_x,m_s^{lh})+m_d^{hh} T_0(E_x,m_s^{hh})]$$

$$\times (q|V_{OX}|-E_g-E_x) \qquad (A.23)$$

(A.21) is for conduction band electron tunneling from <100> silicon. The density-of-state mass and normal mass are $m_d^t=0.417m_0$ and $m_s^t=0.190m_0$ for the 4-fold degenerate transverse valleys and $m_d^l=0.190m_0$ and m_s^l for the 2-fold degenerate longitudinal valleys. For valence band electron tunneling formula (A.22) and valence band hole tunneling formula (A.23), the density-of-state mass and normal mass are

$m_d^{lh}=0.16m_0$ and $m_s^{lh}=0.20m_0$ for the heavy-hole band, and $m_d^{hh}=0.45m_0$ and $m_s^{hh}=0.28m_0$ for the light hole band. The effective mass inside SiO_2 barrier is $m_{ox}=0.437m_0$. E_{FG} and E_{CG} represents the Fermi-energy and conduction band minimum respectively in the bulk of silicon gate material, while the same symbols with subscript 'X' instead of 'G' are used for those in the bulk of silicon substrate materials. V_{SG}, V_{OX} and V_{SX} denote the voltage drops in the silicon gate, SiO_2 insulator and silicon substrate respectively. Electron and hole barrier height and silicon indirect energy gap are $\phi_B=3.13eV$, $\phi_{Bp}=4.25eV$ and $E_g=1.12eV$, which add up to 8.5eV for the SiO_2 energy gap.

The specific forms of $\upsilon(E_x)$ and $T_0(E_x,m_s)$ depends on the barrier shape which are listed as follows:

Trapezoidal Barrier

$$\upsilon = \frac{2\sqrt{2m_{ox}}}{3q\hbar E_{OX}} \left[\phi_B'^{2/3}-(\phi_B'-q|V_{OX}|)^{2/3}\right] \tag{A.24}$$

$$T_0(E_x,m_s) = 16\times \left[\sqrt{\frac{m_s E_x}{m_{ox}\phi_B'}} + \sqrt{\frac{m_{ox}\phi_B'}{m_s E_x}}\right]^{-1} \times \left[\sqrt{\frac{m_s E_x'}{m_{ox}(\phi_B'-q|V_{OX}|)}} + \sqrt{\frac{m_{ox}(\phi_B'-q|V_{OX}|)}{m_s E_x'}}\right]^{-1} \tag{A.25}$$

Triangular Barrier

$$\upsilon = \frac{2\sqrt{2m_{ox}}}{3q\hbar E_{OX}} \phi_B'^{2/3} \tag{A.26}$$

$$T_0(E_x, m_s) = 4\frac{m_s}{m_{ox}} \times \left\{ \sqrt{\frac{m_s E_x}{m_{ox}\phi_B'}} + \sqrt{\frac{m_{ox}\phi_B'}{m_s E_x}} \right\}^{-1} \qquad (A.27)$$

where

$$\begin{aligned} \phi_B' &= \phi_{Bn} - E_x && (J_N) \\ &= \phi_{Bp} - E_x && (J_P) \\ &= \phi_{Bn} + E_x + E_g && (J_{VE}) \end{aligned} \qquad (A.28)$$

$$\begin{aligned} E_x' &= E_x + q|V_{OX}| && (J_N,\ J_P) \\ &= q|V_{OX}| - E_x - E_g && (J_{VE}) \end{aligned} \qquad (A.29)$$

ϕ_B' is the tunneling barrier height for a carrier in the right-hand ($x<0$) initial state with kinetic energy E_x, and E_x' is the carrier kinetic energy in the left-hand ($x>X_{OX}$) final state.

REFERENCES

[1] C. T. Sah, "Evolution of the MOS transistor - from conception to VLSI," Proc. IEEE, 76(10), pp.1280-1326, Oct. 1988. See the history of stabilization of silicon surface on page 1290, and hydrogen annealing of interface traps on pp.1299-1301.

[2] M.M. Atalla, M. Tannerbaum, and E.J. Scheibner, "Stabilization of silicon surface by thermally grown oxides," Bell. Syst. Tech. J., 38(3), p.123, May 1959.

[3] Chih-Tang Sah, "A new semiconductor tetrode, the surface potential controlled transistor," Proc. IRE, vol. 49, no. 11, pp.1623-1634, Nov. 1961.

[4] Chih-Tang Sah, "Effects of surface recombination and channel on p-n junction and transistor characteristics," IRE Trans. Electron Devices, vol. ED-9, no. 1, pp. 94-108, Jan. 1962.

[5] Arnost Neugroschel and Chih-Tang Sah, K. Michael Han, M. S. Carroll, T. Nishida, Jack T. Kavalieros and Yi Lu, "Direct current measurements of oxide and interface traps on oxidized silicon," IEEE Trans. Electron Dev., vol. 42, no. 9, pp. 1657-1662, Sept. 1995.

[6] Chih-Tang Sah, "Reliability of transistors with low-K materials," invited paper, Advanced Metallization and Interconnect Systems for ULSI Applications Conf., Oct. 1997, San Diego. Materials Research Society Proceedings, Mar. 1998, pp. 301-308.

[7] Arnost Neugroschel and Chih-Tang Sah, "Interconnect and MOS transistor degradation at high current densities," Proc. International Reliability Physics Symposium, pp.30-36, March 1999.

[8] C.-T. Sah, "DCIV monitor for diagnosis of sub-quarter-micron technology," report No. C98333, Semiconductor Research Corp., 30pp, Oct. 1998.

[9] C.-T. Sah, "DCIV method does work for thin tunneling gate oxides," report No. C99428, Semiconductor Research Corp., 18pp, Nov. 1999.

[10] C.-T. Sah, "DCIV diagnosis of a tunneling thin gate oxide CMOS technology," report No. C99429, Semiconductor Research Corp., 50pp, Nov. 1999.

[11] Y. Taur, D. A. Buchanan, W. Chen, D.J. Frank, K.E. Ismail, S.-H. Lo, G. A. Sai-Halasz, R.G. Viswanathan, H.-J. C. Wann, S.J. Wind and H.-S. Wong, "CMOS scaling into the nanometer regime," Proceedings of IEEE, 85(4), pp.486-504, Apr. 1997.

[12] R.D. Isaac, "The future of CMOS technology," IBM J. Res. Develop., 44(3), pp.369-378, May 2000.

[13] C.-T. Sah, "Gate tunneling current diagnosis of very thin oxides," report No. C99431, 77pp, and "VFYV measurements of very thin gate oxides," report No. C99430, 35pp, Semiconductor Research Corp., Nov. 1999.

[14] C.-T. Sah, "Insulating layers on silicon substrate," Properties of Silicon, EMIS Datareviews series No.4, section 17.1-17.4, pp. 497-531, INSPEC, IEE, 1988.

[15] C.-T. Sah, Fundamentals of Solid-State Eletronics-Solution Manual, section 912, pp.107-108, World Scientific-Singapore, 1996.

[16] R.B. Laughlin and J.D. Joannopoulos, "Theory of the electronic structure of the Si-SiO_2 interface," Phys. Rev. B21(12), pp.5733-5744, June 1980.

[17] T. Sakurai and T. Sugano, "Theory of continuously distributed trap states at Si-SiO_2 interfaces," J. Appl. Phys., 52(4), pp.2889-2896, April 1981.

[18] Y. Nishi, "Study of silicon-silicon dioxide structure by electron spin resonance," Jpn. J. Appl. Phys., 10(1), pp.52-62, Jan. 1971.

[19] E.H. Poindexter, P.J. Caplan, B.E. Deal, and R.R. Razouk, "Interface states and electron spin resonance centers in thermally oxidized (111) and (100) silicon wafers," J. Appl. Phys., 52(2), pp.879-884, Feb. 1981.

[20] E.H. Poindexter, G.J. Gerardi, M.-E. Rueckel, P.J. Caplan, N.M. Johnson, and D.K. Biegelsen, "Electronic traps and P_b centers at the Si/SiO_2 interface: band-gap energy distribution," J. Appl. Phys., 56(10), pp.2844-2849, Nov. 1984.

[21] G. J. Geraldi, E.H. Poindexter, P.J. Caplan, and N.M. Johnson, "Interface traps and P_b centers in oxidized (100) silicon wafers," Appl. Phys. Lett., 49(6), pp.348-350, Aug. 1986.

[22] J.H. Stathis and E. Cartier, "Atomic hydrogen reactions with Pb centers at the (100) Si/SiO_2 interface," Phys. Rev. Lett., 72(17), pp.2745-2748, April 1994.

[23] J.H. Stathis, "Dissociation kinetics of hydrogen-passivated (100) Si/SiO_2 interface defects," J. Appl. Phys., 77(12), pp.6205-6207, June 1995.

[24] J.T. Krick, and P.M. Lenahan, "Direct observation of interfacial point defects generated by channel hot hole injection in n-channel metal oxide silicon field effect transistors," Appl. Phys. Lett., 59(26), pp.3437-3439, Dec. 1991.

[25] J.H. Stathis and D.J. DiMaria, "Identification of an interface defect generated by hot electrons in SiO_2," Appl. Phys. Lett., 61(24), pp.2887-2889, Dec. 1992.

[26] A. Stesmans, B. Nouwen and V.V. Afanas'ev, "P_{b1} interface defect in thermal (100) Si/SiO_2: ^{29}Si hyperfine interactions," Phys. Rev. B58(23), pp.15801-15809, Dec. 1998.

[27] C.-T. Sah, J.Y. Sun, and J.J. Tzou, "Generation-annealing kinetics of interface states on oxidized silicon activated by 10.2eV photohole injection," J. Appl. Phys., 53(12), pp.8886-8893, Dec. 1982.

[28] L.M. Terman, "An investigation of surface states at a silicon/silicon oxide interface employing metal-oxide-silicon diodes," Solid-State Electron., v5, pp.285-299, Sept. 1962.

[29] C.-T. Sah, A.B. Tole, R.F. Pierret, "Error analysis of surface state density determination using the MOS capacitance method," Solid-State Electron., v12, pp.689-709, Sept. 1969.

[30] E. Nicollian and G. Goetzberger, "The $Si-SiO_2$ interface electrical properties as determined by the metal-insulator-silicon conductance technique," Bell Syst. Tech. J., v46, p.1055, 1967.

[31] G. Groeseneken, H.E. Maes, N. Beltran and R.F. Dekeepsmaecker," A reliable approach to charge-pumping measurements in MOS transistors," IEEE Trans. Electron Devices, vED-31, pp.42-53, Jan. 1984.

[32] C.-T. Sah, A. Neugroschel and K.M. Han, "Current accelerated channel hot carrier stress of MOS transistors," Electronics Letters, 34(2), pp.217-218, Jan. 1998.

[33] Jin Cai and Chih-Tang Sah, "Monitoring interface traps by DCIV method," IEEE Electron Dev. Lett., vol. 20, pp. 60-63, January, 1999.

[34] Jin Cai and Chih-Tang Sah, "Evidence of discrete interface traps on thermally grown thin silicon oxide films," Appl. Phys. Letts, vol. 74, No.2, pp. 257-259, Jan 11, 1999.

[35] K. Huang and A. Rhys, "Theory of light absorption and non-radiative transition in F-centers," Proc. R. Soc. A204, pp.406-423, Dec. 1950.

[36] K. Huang, "Adiabatic approximation theory and static coupling theory of nonradiative transition," Scientia Sinica, v24, pp.27-34, Jan. 1981.

[37] B.K. Ridley, Quantum Processes in Semiconductors, 2nd edition, Oxford University Press, Oxford, 1988. See page 1-2 for the adiabatic approximation.

[38] M.-F. Li, Modern Semiconductor Quantum Physics, World Scientific-Singapore, 1994. See section 390 for lattice relaxation and Jahn-Teller effects.

[39] W. Shockley and W.T. Read, "Statistics of recombination of holes and electrons," Phys. Rev. 87(9), pp.835-842, Sept. 1952. Also, R.N. Hall. "Germanium rectifier characteristics," Phys. Rev. 83, p.228, 1951.

[40] Chih-Tang Sah, Robert N. Noyce and William Shockley, "Carrier generation and recombination in p-n junction of p-n junction characteristics," Proc. IRE, vol. 45, pp. 1228-1243, Sept. 1957.

[41] Chih-Tang Sah, Fundamentals of Solid State Electronics-Study Guide, World Scientific, Singapore, 1993. The exact solutions of voltage equation for the MOS capacitor in fast (1-iteration) convergent forms for the three ranges (inversion, depletion, accumulation) used in this paper are listed on page 129.

[42] Chih-Tang Sah, Transistor Reliability, in Fundamentals of Solid State Electronics-Solution Manual, World Scientific, Singapore, 1996. See section 916, pp. 122-128, for the bond-breaking theory and Table 915.2 on p. 121 for the silicon-hydrogen and silicon-impurity-hydrogen bond energies.

[43] C.-T. Sah, A. Neugroschel, K.M. Han and J.T. Kavalieros, "Profiling interface traps in MOS transistors by the DC current-voltage method," IEEE Electr. Dev. Lett., 17(2), 72-74, Feb. 1996.

[44] B. B. Jie, M. F. Li, C. L. Lou, W. K. Chim, D. S. H. Chan and K. F. Lo, "Investigation of interface traps in LDD pMOST's by the DCIV method," IEEE Electron Device Letters, 15(12), 583-585, 12 November 1997.

[45] C.-T. Sah, "Effects of electrons and holes on the transition layer characteristics of linearly-graded p-n junctions," Proc.IRE, v49(3), 603-618, March 1961.

[46] S. V, Walstra and C.-T.Sah, "Thin oxide thickness extrapolation from capacitance-voltage measurements," IEEE Trans. ED-44(7), 701-707, July 1997.

[47] W. Chen, A. Balasinski and T.-P. Ma, "Lateral profiling of oxide charge and interface traps near MOSFET junctions," IEEE Trans. ED-40(1), 187-196, Jan. 1993.

[48] Bin-Bin Jie, Kok-Hooi Ng, Ming-Fu Li and Keng-Foo Lo, "Correlation between charge pumping method and direct-current current voltage method in p-type metal-oxide-semiconductor field-effect transistors," Japan J. Applied Physics 38(8), Part 1, 4696-4698, August 1999.

[49] K. Michael Han and C.-T. Sah, "Positive oxide charge from hot hole injection during channel-hot-electron stress," IEEE Trans. ED-45(7), 761-763, July 1998.

[50] H. C. Mogul, L. Cong, R. M. Wallace, P. J. Chen, T. A. Rost and K. Harvey, "Electrical and physical characterization of deuterium sinter on submicron devices," Appl. Phys. Letts. 72(14), 1721-1723, April 6, 1998.

[51] Hao Guan, Yaohui Zhang, B. B. Jie, Y. D. He, Ming-Fu Li, Zhong Dong, Joseph Xie, J. L. F. Wang, Andrew C. Yen, George T. T. Sheng and Weidan Li, "Nondestructive DCIV method to evaluate plasma charging damage in ultrathin gate oxide," IEEE Electron Device Letters 20(5), 238-240, May 1999.

[52] R. H. Fowler and L. Nordheim, Proc. Roy. Soc. Lond., Ser. A 119, p.173, 1928.

[53] R.H. Good jr. and E. W. Muller, "Field emission," Handbuck der Physik, Vol. XXI, pp.176-231, Springer-Verlag, Berlin, 1956.

[54] R. H. Fowler, Statistical Mechanics, the Theory of the Properties of Matter in Equilibrium, Chapter XI, 2nd edition, Cambridge Univ. press, 1936.

[55] C. Zener, "A theory of the electrical breakdown of solid dielectrics," Proc. Roy. Soc. Lond., A145, pp.523-529, July, 1934.

[56] L. Esaki, "New phenomenon in narrow germanium p-n junctions," Phys. Rev. 109, pp.603-604, 1957.

[57] E. O. Kane, "Theory of tunneling," J. Appl. Phys., 32(1), pp.83-91, Jan. 1961.

[58] C. T. Sah, "Electronic processes and excess currents in gold-doped narrow silicon junctions," Phys. Rev. 123(5), pp.1594-1621, Sept. 1961.

[59] J. Bardeen, "Tunneling from a many-particle point of view," Phys. Rev. Letters, 6(2), pp.57-59, Jan. 1961.

[60] W. A. Harrison, "Tunneling from an independent-particle point of view," Phys. Rev. 123(1), pp.85-89, July 1961.

[61] J. Shewchun and V. A. K. Temple, "Theoretical tunneling characteristics of the SIS (semiconductor-insulator-semiconductor) diode," J. Appl. Phys., 43(12), pp. 5051-5061, Dec. 1972.

[62] J. G. Simmons, "Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film, "J. Appl. Phys., 34(6), pp.1793-1803, June 1963.

[63] L. L. Chang, P. J. Stiles, and L. Esaki, "Electron tunneling between a metal and a semiconductor: characteristics of Al-Al_2O_3-SnTe and -GeTe junctions," J. Appl. Phys., 38(11), pp.4440-4445, Oct. 1967.

[64] K. H. Gundlach, "Zur Berechnung des tunnelstroms durch eine trapezformige potentialstufe," Solid-State Electr., Vol. 9, pp.949-957, 1966.

[65] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2," J. Appl. Phys., 40(1), pp.278-283, Jan. 1969.

[66] Z. A. Weinberg, "Tunneling of electrons from Si into thermally grown SiO_2," Solid-State Electr., Vol. 20, pp.11-18, Jan. 1977.

[67] Z. A. Weinberg, "On tunneling in metal-oxide-silicon structures," J. Appl. Phys., 53(7), pp.5052-5056, July 1982.

[68] Krieger and Swanson, "Fowler-Nordheim electron tunneling in thin Si-SiO_2-Al structures," J. Appl. Phys., 52(9), pp.5710-5717, Sept. 1981.

[69] G. Lewicki and L. Maserjian, "Oscillations in MOS tunneling," J. Appl. Phys., 46(7), pp.3032-3039, July 1975.

[70] M. V. Fischetti, D. J. DiMaria, L. Dori, J. Batey, E. Tierney,a nd J. Stasiak, "Ballistic electron transport in thin SiO_2 dioxide films," Phys. Rev. B, 35(9), pp.4404-4415, Mar. 1987.

[71] S. Zafar, K. A. Conrad, Q. Liu, E. A. Irene, G. Hames, R. Kuehn, and J. J. Wortman, "Thickness and effective electron mass measurements for thin silicon dioxide films using tunneling current oscillations," Appl. Phys. Lett., 67(7), pp.1031-1033, Aug. 1995.

[72] Z.A. Weinberg, W.C. Johnson, and M.A. Lampert, "Determination of the sign of carrier transported across SiO_2 films on Si," Appl. Phys. Lett., 25(1), pp.42-43, July 1974

[73] Z.A. Weinberg and R.A. Pollak, "Hole conduction and valence-band structure of Si_3N_4 films," Appl. Phys. Lett., 27(4), pp.254-255, Aug. 1975.

[74] L.D. Yau, "Arguments for electron conduction in silicon nitride," IEEE Electr. Dev. Lett., 5(8), pp.318-321, Aug. 1984.

[75] B. Eitan and A. Kolodny, "Two components of tunneling current in metal-oxide-semiconductor structures," Appl. Phys. Lett., 43(1), pp.106-108, July 1983.

[76] Z. A. Weinberg and M. V. Fischetti, "Investigation of the SiO_2-induced substrate current in silicon field-effect transistors," J. Appl. Phys., 57(2), pp.443-452, Jan. 1985.

[77] M. Rasras, I.De Wolfe, G. Groeseneken, B. Kaczer, R. Degraeve, and H.E. Maes, "Photo-carrier generation as the origin of Fowler-Nordheim-induced substrate hole current in thin oxides," IEDM 18.8, pp.465-648, Dec. 1999.

[78] A. Shanware, J.P. Shiely, H.Z. Massound, E. Vogel, K. Henson, A. Srivastava, C. Osburn, J.R. Hauser and J.J. Wortman, "Extraction of the gate oxide thickness of n- and p-channel MOSFETs below 20Å from the substrate current resulting from valence-band electron tunneling," IEDM, 34.1, pp.815-818, Dec. 1999.

[79] C.-T. Sah, Fundamentals of Solid-State Electronics, World Scientific, Singapore, 1991. See page 356 for electron affinity data: 4.029eV for Si and 0.9eV for SiO_2 which gives a Si/SiO_2 conduction band offset of 3.13eV. Valence band offset is derived from energy gaps of Si (1.12eV) and SiO2 (8.5eV) also listed on p.356.

[80] Y. Lu and C.-T. Sah, "Energy momentum conservation during energetic carrier generation and recombination in silicon," Phys. Rev. B52(8), pp.5657-5664, Aug. 1995.

[81] Y. Lu and C.-T. Sah, "Two pathways of positive oxide-charge buildup during electron tunneling into silicon dioxide film," J. Appl. Phys., 76(8), pp.4724-4727, Oct. 1994.

[82] T. Matsuoka, S. Kakimoto, M. Nakano, H. Kotaki, S. Hayashida, K. Sugimoto, K. Adachi, S. Morishita, K. Uda, Y. Sato, M. Yamanaka, T. Ogura and J. Takagi, "Direct tunneling N_2O gate oxynitrides for low-voltage operation of dual gate CMOSFETs," IEDM 34.2, pp.851-854, Dec. 1995.

[83] Y. Shi, T.P. Ma, S. Prasad and S. Dhanda, "Polarity dependent gate tunneling currents in dual-gate CMOSFET's," IEEE Trans. Electr. Dev., 45(11), pp.2355-2360, Nov. 1998.

[84] W.-C. Lee, T.-J. King, and C. Hu, "Evidence of hole direct tunneling through ultrathin gate oxide using p+ poly-SiGe gate," IEEE Electr. Dev. Lett., 20(6), pp.268-270, June, 1999.

[85] R.H. Dennard, F.H. Gaensslen, H.-N. Yu, V.L. Rideout, E. Bassous,a nd A.R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," IEEE J. Solid-State Circuits, SC-9(5), pp.256-268, Oct. 1974.

[86] H.S. Momose, M. Ono, T. Yoshitomi, T. Ohguro, S.-I. Nakamura, M. Saito and H. Iwai, "1.5nm direct tunneling gate oxide Si MOSFET's," IEEE Trans. Electr. Dev., 43(8), pp.1233-, Aug. 1996.

[87] M.S. Krishnan, L. Chang, T.-J King, J. Bokor,a nd C. Hu, "MOSFETs with 9 to 13 thick gate oxide," IEDM 10.1, pp.241-244, Dec. 1999.

[88] K. Ahmed, E. Ibok, G. C.-F. Yeap, Q. Xiang, B. Ogle, J.J. Wortman, and J.R. Hauser, "Impact of tunnel currents and channel resistance on the characterization of channel inversion layer charge and polysilicon-gate depletion of sub-20-A gate oxide MOSFETs," IEEE Trans. Electr. Dev., 46(8), pp.1650-1655, Aug. 1999.

[89] S.-H. Lo, D.A. Buchanan, Y. Taur and W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's," IEEE Electr. Dev. Lett., 18(5), pp.209-211, May 1997.

[90] S.-H. Lo, D.A. Buchanan and Y. Taur, "Modeling, characterization of quantization, polysilicon depletion and direct tunneling effects in MOSFETs with ultrathin oxides," IBM J. Res. Develop., 43(3), pp.327-337, May 1999.

[91] W-K. Shih, E.X. Wang, S. Jallepalli, F. Leon, Christine M. Maziar and A. F. Tasch, "Modeling gate leakage current in nMOS structures due to tunneling through an ultra-thin oxide," Solid-State Electr., 42(6), pp.997-1006, June 1998.

[92] N. Yang, W.K. Henson, J.R. Hauser, and J.J. Wortman, "Modeling study of ultra-thin gate oxides using direct tunneling current and capacitance-voltage measurements in MOS devices," IEEE Trans. Electr. Dev., 46(7), pp.1464-1471, July 1999.

[93] F. Rana, S. Tiwari and D.A. Buchanan, "Self-consistent modeling of accumulation layers and tunneling currents through very thin oxides," Appl. Phys. Lett., 69(8), pp.1104-1106, Aug. 1996.

[94] L.F. Register, E. Rosenbaum and K. Yang, "Analytic model for direct tunneling in polycrystalline silicon-gate metal-oxide-semiconductor devices," Appl. Phys. Lett., 74(3), pp.457-459, Jan. 1999.

[95] J.J. Sakurai, Modern Quantum Mechanics, chapter 2, pp.104-106, Addison-Wesley, 1994.

[96] Y.P. Li and W.Y. Ching, "Band structures of all polycrystalline forms of silicon dioxide," Phys. Rev. B, 31(4), pp.2172-2179, Feb. 1985.

[97] L.B. Freeman and W.E. Dahlke, "Theory of tunneling into interface states," Solid-State Electr., 13(11), pp.1483-1503, Nov. 1970.

[98] F. Stern, "Self-consistent results for n-type Si inversion layers," Phys. Rev. B, 5(12), pp.4891-4899, June, 1972.

[99] Class notes of EEL7936, taught by Professor C.-T. Sah at Univ. Florida in the Fall semester of 1999.

[100] The National Technology Roadmap for Semiconductors Technology Needs, pp.71-72, Semiconductor Industry Association, 1997.

[101] C.-T. Sah, Fundamentals of Solid-State Electronics, Singapore: World Scientific, 1991. See page 490 for the experimental work functions of the metals in vacuum and also table 413.1 on page 356.

[102] The National Technology Roadmap for Semiconductors Technology Needs, Semiconductor Industry Association, 1997 edition. See page 74, table 22.

[103] T.Y. Chan, J. Chen, P.K. Ko and C. Hu, "The impact of gate-induced drain leakage current on MOSFET scaling," IEDM Technical Digest, pp.718-7121, 1987.

[104] J. J. Quinn, G. Kawamoto and B.D. McCombe, "Surface spectroscopy by surface channel tunneling," Surface Sci., Vol. 73, pp.190-196, 1978.

[105] L.V. Keldysh, "Behavior of non-metallic crystals in strong electric fields," Soviet Phys. JETP, 6(33), pp.763-770, Apr. 1958.

[106] P.J. Price and J.M. Radcliffe, "Esaki Tunneling," IBM J. Research Develop., 3(4), pp.364-371, Oct. 1959.

[107] C.B. Duke, Tunneling in Solids, Academic Press, New York and London, 1969. See Chapter 4, section 7 and 8.

[108] E.O. Kane, "Zener tunneling in semiconductors," Phys. Chem. Solids, Vol. 12, pp.181-188, Dec. 1959.

[109] J. Maserjian, "Tunneling in thin MOS structures," J. Vac. Sci. Technol., Vol. 11, pp.996-1003, 1974.

[110] J.C. Hensel, and G. Feher, "Cyclotron resonance experiments in uniaxially stressed silicon: valence band inverse mass parameters and deformation potential," Phys. Rev., Vol.129, pp.1041-1062, Feb. 1963.

[111] J.L. Moll, Physics of semiconductors, McGraw-Hill, 1964. See Ch. 12, p.250.

[112] J.R. Schrieffer, "Mobility in inversion layers: theory and experiment," Semiconductor Surface Physics, Univ. Penn. Press, 1957. See p.68.

[113] T.Ando, A.B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems," Rev. Modern Phys., 54(2), pp.437-672, April 1982.

[114] T. Ando, "Density-functional calculations of sub-band structure in accumulation and inversion layers," Phys. Rev. B13(8), pp. 3468-3477, April, 1976.

[115] J. Sune, P. Olivo and B. Ricco, "Self-consistent solution of the Poisson and Schrodinger equations in accumulated semiconductor-insulator interfaces," J. Appl. Physics, 70(1), pp.337-345, July, 1991.

[116] J. Cai and C.-T. Sah, "DCIV diagnosis monitor for submicron MOS transistor designs and fabrication processes," SRC Techcon paper 5.8, Sept., 1998.

[117] A. Alcovic, M. Dutoit and M. Ilegems, "Characterization of hot-electron stressed MOSFET's by low-temperature measurements of the drain tunnel current," IEEE Trans. Elect. Dev., 37(6), pp.1467-1476, June, 1990. See Fig. 3.

BIOGRAPHICAL SKETCH

Jin Cai received the B.S. degree in physics from Fudan University, Shanghai, in 1989, and the M.S. degree in electrical engineering in 1997 from the University of Florida, Gainesville. He will receive the Ph.D. degree in electrical engineering from the University of Florida in 2000 under the guidance of Professor Chih-Tang Sah. His Ph.D. dissertation research concerns the direct current measurement of interface traps using the DCIV method, and its application as a monitor for deep submicron silicon transistor design and processing.

During the summer of 1999, he interned at IBM Research Laboratories in Yorktown Heights, where he characterized sub-50nm CMOS transistors from 77K to 300K. From 1989 to 1995, he was employed by the Philips Semiconductor, Shanghai. He was assigned, during the first year, to Philips' Nijmegen facility in the Netherlands for training in silicon process technologies. He was then appointed the engineering section head in photolithography to start-up Shanghai-Philips' bipolar and CMOS manufacturing facility. He has published four articles on oxide and interface traps and the DCIV diagnosis monitors in IEEE and AIP journals.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Chih-Tang Sah, Chairman
Robert C. Pittman Eminent Scholar and
Graduate Research Professor of
Electrical and Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Arnost Neugroschel
Professor of Electrical and Computer
Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Ewen M. Thomson
Associate Professor of Electrical and
Computer Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Sergei Obukhov
Associate Professor of Physics

This thesis was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirement for the degree of Doctor of Philosophy.

August 2000

M. Jack Ohanian
Dean, College of Engineering

Winfred M. Phillips
Dean, Graduate School

937 F8 T 607
10/16/00 34760 SELF
INFORMATION CONSERVATION INC.